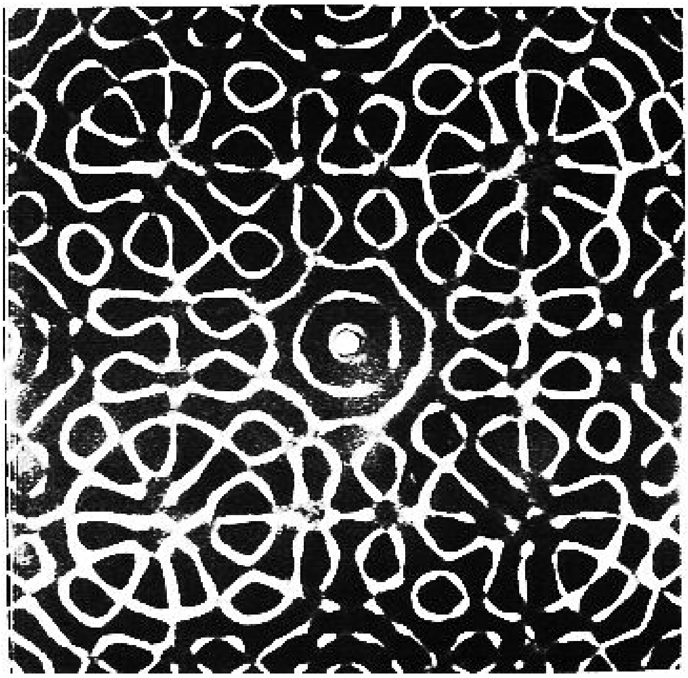
CHALMERS
UNIVERSITY OF TECHNOLOGY



Investigation on the Effect of Loudspeaker Ringing on Perceived Spectral Balance

Master's thesis in Sound And Vibration

VASISHTA KANTHI

Department of Civil And Environmental Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Master's thesis 2018: ACEX30-18-112

Investigation on the Effect of Loudspeaker Ringing on Perceived Spectral Balance

VASISHTA KANTHI

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology
Gothenburg, Sweden 2019

Investigation on the Effect of Loudspeaker Ringing on Perceived Spectral Balance

Supervisor: Jens Ahrens, Division of Applied Acoustics
Examiner: Jens Ahrens, Division of Applied Acoustics

Master's Thesis 2018: ACEX30-18-112
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Chladni Plate Resonance - A visualization of Resonances in a plate

Typeset in LaTeX
Printed by [Name of printing company]
Gothenburg, Sweden 2019

Investigation on the Effect of Loudspeaker Ringing on Perceived Spectral Balance
VASISHTA KANTHIMATHINATHAN SUBRAMANIAN
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

The Ringing artefacts of a loudspeaker can have a large impact on the perception of music in home, studio and car audio systems. A large contributing factor to this ringing effect is given by the driver's natural resonance. In the case of car audio systems, additional mounting components and cavities within a car's door and body of the driver can enhance the ringing effect. Unfortunately, the ringing effect cannot be eliminated completely, as it is an intrinsic property consequent to any resonating object. Alternatively, its effect can be controlled. A thorough investigation on different methodologies for detection and measurement of the driver's natural resonances, as well as the perceptual effect of ringing caused and enhanced by the resonances from cavities, mounting enclosures and other components of a car door was implemented. A comparison of different systems, created with synthetic transfer functions was performed, in which artificial low frequency resonances were employed. These artificial low frequency resonances have similar characteristic properties of the driver resonances, in which they are detected and measured by using signal processing techniques involving Cumulative Spectral Decay (CSD) and Continuous Wavelet Transform (CWT) plots, as well as a System Identification Method that uses the Steiglitz - McBride algorithm. These methods rely on the impulse response measurement of a loudspeaker. The methodologies successfully detected the driver's natural resonance frequency and its strength in terms of Quality (Q) factor, with a minimal degree of variance from each other. From these results, three low frequencies and their corresponding Q factors were chosen, and were set into a listening test platform as second order IIR peak filters. This was implemented by using MATLAB's GUI and Simulink interface. The test used two music audio of different genres: a recorded jazz ensemble and a synthetic electronic ensemble. To emulate the effect of delayed resonances, three delay time values were given into the resonance settings, and the test was carried out. The test primarily focuses on three case studies: the threshold of audibility of resonances, the threshold at which a non resonant system with a parametric bandpass filter is perceptually similar to a resonant system, and the threshold at which a system's resonance is inaudible when the same filter is used. The results show a high degree of dependency between the frequency in test and the audio. Very low frequencies are found to have the highest audibility thresholds in the case of recorded audio. The control of ringing is dependent on the type of audio played, as indicated by the high variance in the results with the recorded audio in comparison to the synthetic audio. The effect of delayed resonances has had a minimal impact on the results. This shows a good promise, and further future work is definitely needed in order to create a metric for loudness of a mounted car loudspeaker resonance, as this metric can intuitively suggest a severity threshold to the effect of ringing in the spectral balance of audio. This thesis work serves as a starting point into developing the metric for ringing loudness.

Keywords: Ringing, Resonance, Quality Factor, Centre Frequency, CSD, Wavelet, System Resonance, Listening Test

Acknowledgements

I would like to thank my Supervisor, Jens Ahrens, who has been the backbone of my progress and inspiration to reach my intended goal in my current field of interest. Thank you for all the patience, time and energy spent in carrying out this project successfully, and at the same time, giving me all the knowledge needed in the field, which I will take forward for years to come, and teach others as well.

I am very thankful and proud of my classmate and friend, Nikolaos Chrysovalantis Roumpakis, who not only discussed on critical problems and issues in great detail, but also has supported me morally and personally during the course of the thesis and studies.

I thank all the staff and students at the Division of Applied Acoustics at Chalmers University of Technology, who have been providing a friendly and motivating environment.

Finally, I thank my parents and grandparents. Thank you for all the moral and emotional support throughout my years in life, through my times of despair and happiness, which has honed me to be the man today

Vasishta Kanthimathinathan Subramanian, Gothenburg, May 2019

Contents

List of Figures

List of Tables

1
Introduction

Music, in its most intrinsic form, is the art of combining tones or sounds in order, created by instruments of different forms, that create a certain or specific composition, that creates harmony and expression of emotion. These instruments may be vocal, percussive, strings, air flow, or even electronic. Each of these instruments have a unique characteristic in tonality, timbre, colour and quality, depending on its assembly and function to produce the sound, that specifically defines the role of the instrument in a certain ensemble. The main characteristic in question, is resonance. Their resonant behaviour describes the type of tonality, timbre, colour and quality the instrument may have, and its purpose in a composition. In fact, it is due to the sound of these resonances that many musical genres today have an indirect dependency. For example, the classical sound of Flamenco comes from the classical Spanish guitar. In Indian classical music such as Carnatic, the Saraswathi Veena, Mridangam and Tampura are some unique instruments that have specific resonant behaviours that define the quality of the genre. In Jazz, the incorporation of piano, saxophone, trumpet drums, guitar and contra bass give their unique feature individually, or as a whole. In all types of genres, the human voice is the most common instrument, that has several resonances, such as mouth resonance, nasal resonance, chest resonance, etc., in which each resonance give a very distinctive quality and colour to the tonal sound of the human voice. In essence, resonances are a crucial part in the art form of music. At some level, it dictates the type, the harmony and emotion to its art. However, resonances from other bodies can prove to be hazardous. Any object that can vibrate, will have a resonant frequency in which, theoretically, would cause the object to vibrate indefinitely. As objects are subjected to natural dampness from its environments, its vibrational energy will dissipate over a short period of time. The main difference in this case, is that the decay time of a vibrational object at its resonant frequency is the longest, and this period of longevity causes the problem of "Ringing".

This brings the implication that not every resonance in an acoustical system provides any pleasantness, especially with loudspeaker driver and port-reflex resonances. The effect of these loudspeaker resonances result in the addition of unwanted, sustained sounds that can merge with audio. This is conveniently called ringing noise. Fortunately, these resonances being low frequency resonances that range between 40 - 50 Hz in the case of monitor and mid-range loudspeakers, and 15 - 45 Hz in the case of sub woofers can be minimised. There has been a considerable amount of research for decades on minimising the effect of low frequency driver resonances. Most of today's high end loudspeakers such as Genelec, Neumann, Creative, etc., have a good overall frequency response. Most of this research focus on the design aesthetics of the loudspeaker driver and enclosure, usage of filters, etc., which improve the Thiele - Small Parameters and overall frequency response.

The same cannot be said in the case of car loudspeakers. High end loudspeakers such as Genelec, Neumann, etc., have their own, well designed enclosures and electric circuitry that produce its high quality audio, with an overall flat frequency response. Car loudspeakers, however, have a common enclosure, being the car's internal body. This suggests the problem of having a non-stationary enclosure mount, that could induce vibrations directly into the driver, primarily due to the car movement and secondarily due to a varying compliance. As a result, there will be a shift in the driver's resonance frequency. Modern cars of today have enclosure mountings with some degree of damping, that isolate the effect of vibrations. In spite of this, there is still an eerie of doubt in regard to the audibility of these resonances, especially when audio is played.

Another issue that car loudspeaker systems have is the inclusion of cavities within its mounting. As new car designs come by, the interior design may not take into account the loudspeaker's design aesthetics. As a result, this would induce cavity holes around a driver's mounting. The main danger in this case is the production of cavity resonances. As explained in the beginning, these resonances will have its characteristic perceptual behaviour, and can affect the perceptual spectral balance if the input energy is high. Since the driver can induce high energy vibrations and since the driver's resonant frequency could be shifted due to the varying compliance experienced due to the car's internal enclosure, the probability of inducing audible cavity resonances is high.

Unlike a loudspeaker driver's resonance, it is much harder to control the effect of cavity resonances, as it is implicit to the interior design of the car. Localization of these resonances is complex, as it varies with varying interior designs. Conventional Sound Pressure Level (SPL) measurements are impractical, as it is rather complex to deduce level differences between a resonance and background noise when heavy backward masking effects is experienced. Psychoacoustical models to deduce loudness and roughness of these resonances can also be hard to detect for the same reason.

In spite of all the nuances, fortunately the effect of ringing noise by these resonances can be dealt with. But some questions do arise:

1. How does one quantify the loudness and roughness of ringing? As discussed already, conventional methods are impractical
2. How much control is necessary to minimise the ringing noise effectively? One could use filters or damping methods to minimise the effect, but does that come at the cost of affecting the spectral balance of audio?
3. How much does a Loudspeaker Driver's Resonance contribute to the effect of car-induced resonances, knowing that the primary source of vibrational energy is from the loudspeaker's driver?
4. How can one determine a standard or metric to give a severity value to Ringing? Since car-interior designs change over short periods of time, how does one ensure that the new designs fall within agreeable limits?

1.1 Scope of the Thesis

The uncertainty over the audibility of car-interior-induced resonances through the spectral balance of audio raises the need to have a metric model to detect, estimate and quantify the effect of these resonances. The model should also provide a severity threshold, by estimating the loudness and roughness parameters of the resonances, and correlating them to the car-interior design with its audio system. As this is a rather complex problem that requires a series of tests that pose several challenges in themselves and requires ample amount of time to implement the model, this thesis serves as a starting point into creating the metric model. The thesis will explore the possibility of determining an adequate methodology based on studies such as [1],[2] and [4] to detect and estimate resonance characteristics, and investigate their audibility through a series of listening tests by creating a resonance model. The methodology used to detect and estimate the resonances will be tested with a set of loudspeaker drivers, and the estimated values will provide the range of centre frequencies and quality factor values that can be set on the listening test model. The listening test model will be tested on a number of participants, and a perceptual evaluation of their experiences will be quantified, and investigated.

2
Theory

A major part of the methodologies used to detect and estimate resonances, as well as creating the resonance model involves many signal processing concepts and techniques. In this section, a basic theory of several techniques and methods used to evaluate the perception of loudspeaker ringing is discussed. Most of the theory explained in this section are based on the signal processing theory.

2.1 Impulse Response

The impulse response of a system is defined as the response or reaction of a system when subjected to a pulse. The impulse response defines a system's behaviour to a signal and provides an altered time and frequency response as output.

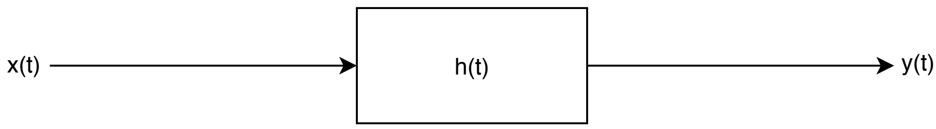


Figure 2.1: Block Diagram of a System with an impulse response of h(t)

From the block diagram in figure 2.1, shows a general representation of a system with an impulse response h(t), with an input signal x(t) and output signal y(t). The impulse response h(t) is generally represented as the ratio between the output signal y(t) and the input signal, x(t).

$$H(s) = \frac{Y(s)}{X(s)} \tag{2.1}$$

where H(s) is the Laplace Transform of h(t). The right side of equation 2.1 gives the Transfer function of a system.

2.2 Convolution

Convolution is an operation of two functions to form a third function, in which it expresses the how the shape of one is modified by the other. This is an important concept in signal processing, as it suggests how a signal's response can be altered by the response of a system.

$$y(t) = x(t) \circledast h(t) \tag{2.2}$$
$$= \int_0^t x(t)h(t-\tau)dt \tag{2.3}$$

In equation 2.2, the envelope of the impulse response h(t) modifies the input signal x(t), which leads to a modified signal y(t).

2.3 Laplace Transform

The Laplace Transform (LT) is an integral function, which takes the function of a real variable and converts it to a function of a complex variable 's'.

$$F(s) = \int_0^\infty f(t)e^{-st}dt \tag{2.4}$$

where 's' is a complex number equal to $\sigma + j\omega$. The formal definition states that the real variable time 't' can be transformed into a complex frequency variable 's'. LT has an intrinsic property, which states that the convolution of two signals is equivalent to the product of their Laplace transforms. Looking at equation 2.2,

$$y(t) = x(t) \circledast h(t) \tag{2.5}$$
$$\mathcal{L}[y(t)] = \mathcal{L}[x(t) \circledast h(t)] \tag{2.6}$$
$$Y(s) = X(s)H(s) \tag{2.7}$$

LT plays an important role in control theory, as it can represent the convolution of a linear time invariant system as a multiplication factor and allows to define the transfer function of a system (as shown in equation 2.1).

2.4 Fourier Transform

The Fourier Transform (FT) of a function is defined as the representation of a real function as the sum of infinite number of periodic complex sine waves. It is an extremely powerful tool in representing the frequency components of a system from its time function, in this case, its impulse response.

$$F(\omega) = \int_{-\infty}^\infty f(t)e^{-j\omega t}dt$$

Like LT, the FT has the intrinsic property of representing the convolution of two signals as the multiple of their Fourier Transforms. Looking at equation 2.2,

$$y(t) = x(t) \circledast h(t) \tag{2.8}$$
$$\mathcal{F}[y(t)] = \mathcal{F}[x(t) \circledast h(t)] \tag{2.9}$$
$$Y(\omega) = X(\omega)H(\omega) \tag{2.10}$$

Equation 2.10 can also represent the transfer function of the system. The main difference between the transfer functions in LT domain and FT domain is that LT can represent the transient response of a system, where as FT can represent the steady state response of a system.
In signal processing, the FT of a signal is implemented by the Discrete Fourier Transform (DFT) of the signal, which is given by the equation:

$$F(n) = \sum_{0}^{N-1} f(k)e^{-j\frac{2\pi}{N}nk} \tag{2.11}$$

where $n = 0 : N - 1$, represents the sample value at an instance ‘k’, equivalent to time. The DFT is a necessary tool, since digital signals are evaluated in samples. A faster process algorithm of DFT is the Fast Fourier Transform (FFT), in which it implements the DFT given in equation 2.11 when N is a power of 2. The result of this is that the number of operations in DFT, which is $O(N^2)$ is reduced to $O(Nlog_2N)$ operations. A resultant of this property in FFT is the Fast Convolution, which is explained in detail in Appendix **??**. It is basically, a faster computation algorithm to the conventional convolution.

2.5 Cumulative Spectral Decay (CSD)

The cumulative spectral decay (CSD) of a system is the time - frequency representation of the system’s impulse response. It basically shows the decay of frequency components of a system over a time scale. It is a form of a Short Time Fourier Transform (STFT), with the difference that it uses a Unit Step Function as a window in conjunction with the impulse response of the system

$$C(\tau, \omega) = \int_{-\infty}^{\infty} f(t)U(t-\tau)e^{-j\omega t}dt \tag{2.12}$$

where ‘$U(t-\tau)$’ is the Unit Step Function at a certain time ‘τ’. In the equation 2.12, the product of f(t) and $U(t-\tau)$ indicates the convolution product between the signal and the unit step function. This suggests that the CSD of a system can be implemented by computing the Fast Fourier Transform of its impulse response at equivalent time slices.

2.6 Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) is a time-frequency representation of a signal which uses an integral function, known as the wavelet, and involves convolution of the integral function with the input signal. This is an alternate method of to STFT, as it’s time-frequency resolution is no longer fixed, but dependent on its ‘scale’ coefficients.

$$W(u, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t)\phi(\frac{t-u}{s})dt \tag{2.13}$$

where '$\phi(\frac{t-u}{s})$' is a wavelet basis function that is orthonormal, 'u' is an instance in time, and 's' is the scaling factor. There are many types of wavelets that satisfy the orthonormal condition, and have distinct properties that have a specific use in many areas. From studies given in [9] and [10], the Morlet Wavelet is very well suited for the purpose of identifying loudspeaker related properties, as well as a conventional tool in decomposing the frequency response from the impulse response.

$$\phi(t) = \frac{1}{\sqrt{\pi B}} e^{j\omega_0 t} e^{\frac{-t^2}{B}} \tag{2.14}$$

where 'ω_0' indicates the centre frequency of the wavelet, and 'B' is the bandwidth parameter, that has a control over the decay of the oscillation in the wavelet [9].

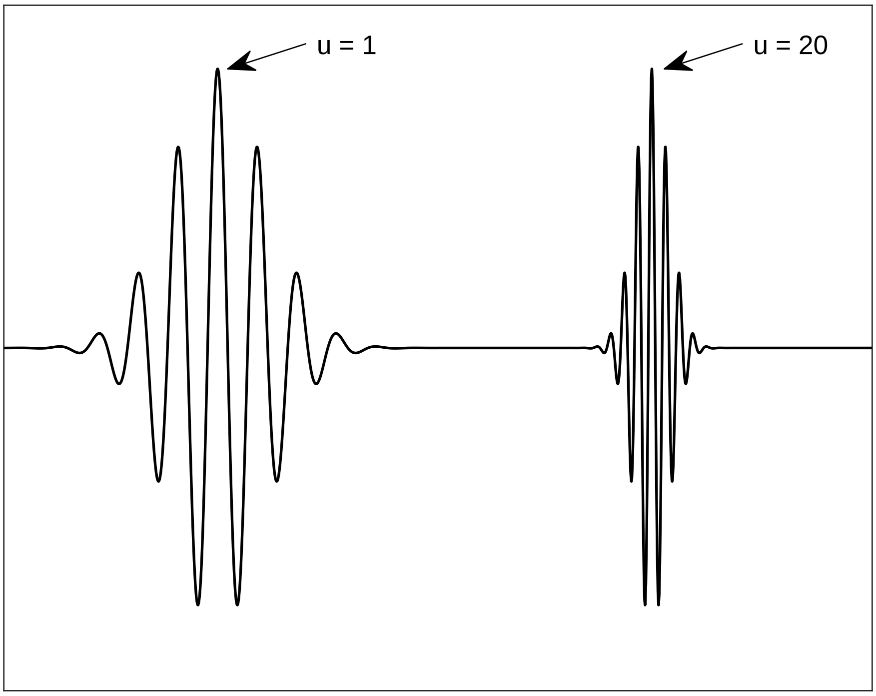


Figure 2.2: Unscaled and Scaled Morlet Wavelet

The scaling factor applied is used to stretch or compress the wavelet, such that it changes the frequency of the wavelet. Figure 2.2 shows the morlet wavelet, when unscaled and scaled. The convolution of the impulse response and these scaled wavelets would obtain a filtered impulse response, representing an approximate frequency, given by the following equation.

$$f = \frac{f_c f_s}{u} \tag{2.15}$$

f_c being the centre frequency of the wavelet, f_s being the sampling frequency of the wavelet, and ‘u’ being the scaling factor.

3
Description

To have a clear conscious on the problem and the scope of the thesis, this chapter describes the definition of Ringing, how it is created, and how it affects the response of a system. It also describes how a perceptual model can be implemented to evaluate the perception of ringing in the spectral balance of audio.

3.1 Ringing

When audio is played through a loudspeaker, if the audio consists of frequencies that match with the resonance frequency of the loudspeaker, the diaphragm will start to vibrate indefinitely[1] and produces its own sound, in which not only will affect the quality of audio, but will also keep vibrating and have a sustained effect, even when the signal has stopped. This is conveniently known as "Ringing". The cause for the loudspeaker ringing effect is due to the fact that the driver is a form of a mechanical mass-spring system, in which the cone attached to the wire coil acts as the mass, and the foam or rubber surrounds that joins the inner and outer edges to the frame acts as the spring[2] [3]. As a mass-spring system, there will be a frequency that causes the diaphragm to vibrate indefinitely. This frequency, being the mechanical resonance frequency of the driver. This also results in a generation of a back electromotive force (EMF) that travels back to the loudspeaker cables and to the power amplifier. Essentially, any vibrating object that has a mass-spring characteristic would definitely have a resonating frequency, thereby having the tendency to induce ringing.

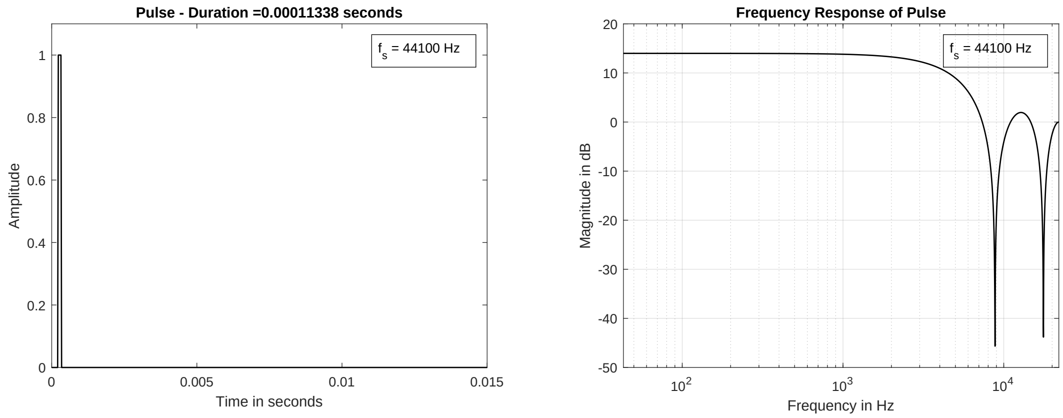
Ringing is a characteristic associated with resonance. The amount of ringing from a resonance depends on the strength and Q of the resonance. In loudspeakers, the ringing effect of the mechanical resonance from the driver can be perceptive, depending on the loudness of the input signal, and the strength of the resonance. Generally, ringing does not add new frequencies, unlike distortion, rather ringing sustains existing frequencies[4]. This may not always be the case, if the strength of the resonance is high. For instance, a room having prominent modes can flatten audio that have frequencies close to the resonant frequency. The amount of flatness is proportional to the strength and Q of the room resonance.

[1] Theoretically, a resonance will vibrate indefinitely. In reality, a resonance will be subjected to damping, in which the energy will dissipate over time

[2] i.e., The diaphragm of the driver

3.2 Ringing Example: Pulse

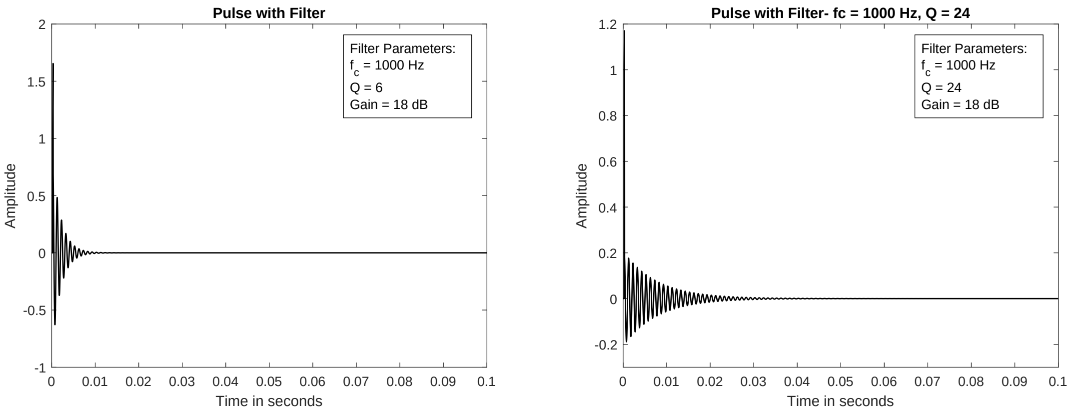
The following example is an illustration of ringing, based on [4]. Consider a signal consisting of a short-duration pulse, created with a sampling frequency, $f_s = 44100\ Hz$, with a short initial delay, as shown in figure 3.1a. This pulse consists of all frequencies, from 0 to 22.5 kHz. This can be clearly seen in figure 3.1b.



(a) Short Duration Pulse **(b)** Frequency Response of the Pulse

Figure 3.1: Time Signal and Frequency Response of a Short Duration Pulse

The signal is passed through two peak filters, in which both have a centre frequency f_c equal to 1000 Hz and gain of 18 dB. The Q factors of both filters are set to 6 and 24. The resultant output is shown in figure 3.2.

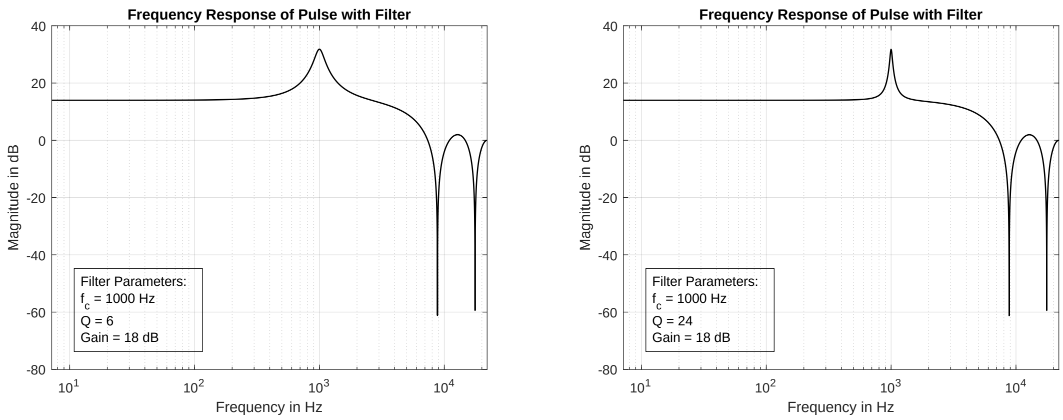


(a) Pulse through filter - Q = 6 **(b)** Pulse through filter - Q = 24

Figure 3.2: Time Signal of Pulse through Filters

The effect of ringing can be clearly seen in both cases. Although this example is seen in the case for a pulse, the same effect also occurs with audio or speech passed through a loudspeaker. When the input signal contains frequencies that align closely to the centre frequency, it will result in a boost, as well as sustain. The amount of sustain is proportional to the Q value. In figure 3.3a, although the Q is seen to be

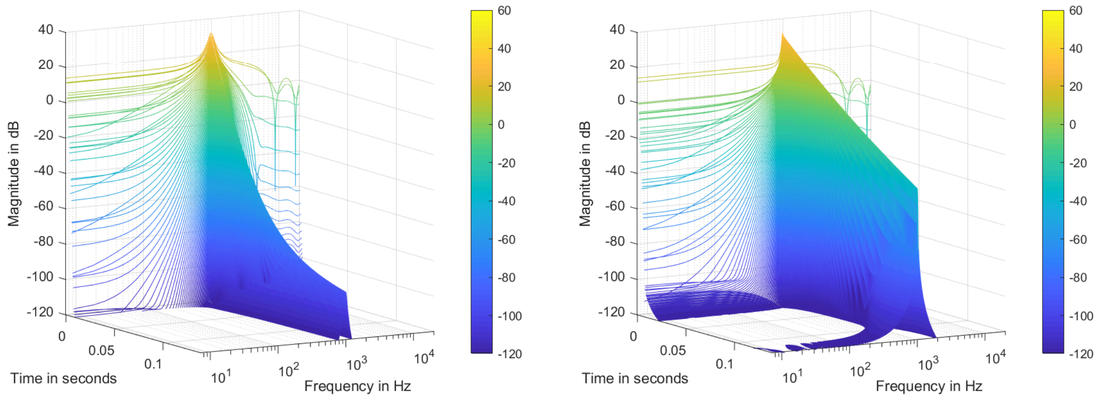
small, the ringing is sustained up to 10 milliseconds. In the case for a Q value of 24 as shown in figure 3.3b, the ringing is sustained up to 40 milliseconds.
As mentioned, the resonance will boost and sustain the frequencies close to its centre frequency. This is shown clearly in figure 3.3.



(a) Q = 6 **(b)** Q = 24

Figure 3.3: Frequency Response of Pulse through Filters

Clearly, the boost at 1000 Hz can be seen. However, there is no information shown on the evolution of frequencies over time. In order to do this, a Short Time Fourier Transform (STFT) is needed to have a time - frequency display. STFT has the time - frequency resolution issue, depending on the type of window used to implement the plot. An alternate approach to the general STFT is the cumulative spectral decay (CSD), which is a form of STFT, in which it uses a unit step function as the window.



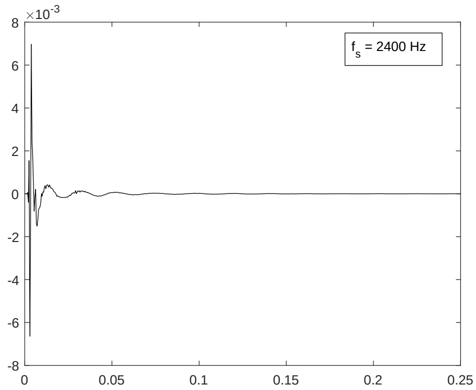
(a) Q = 6 **(b)** Q = 24

Figure 3.4: CSD of a Pulse through Filters

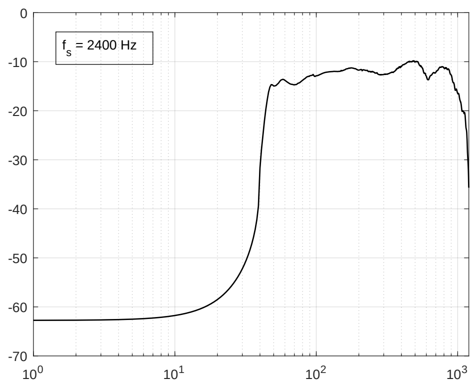
The spectral decay profiles are more prominent, as shown in figure 3.5. The sustain observed in figure 3.2 can now be correlated to the given CSD plots. It is much clear to see how the time taken for the resonance to decay varies with varying Q.

3.3 Ringing Example: Genelec 8020B

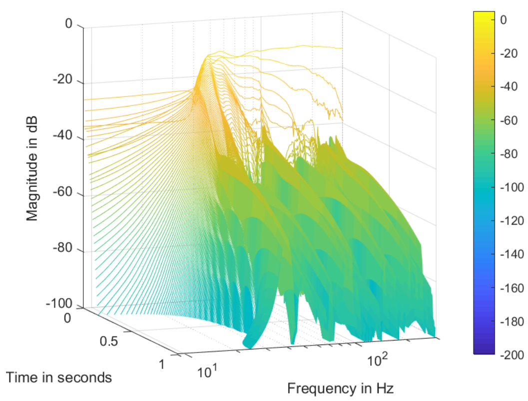
To illustrate the effect of strong resonances with varying Q, an impulse response measurement of a Genelec 8020B monitor Loudspeaker is taken as an example. The measurement procedure is explained in Chapter 4. In consideration to the primary focus of low frequency resonances, the measured impulse response has been down-sampled to a sampling frequency of 2400 Hz, originally being 51200 Hz[3], since high frequency content are of no interest, as well as the necessity for ease in computational efficiency.



(a) Impulse Response



(b) Frequency Response



(c) CSD

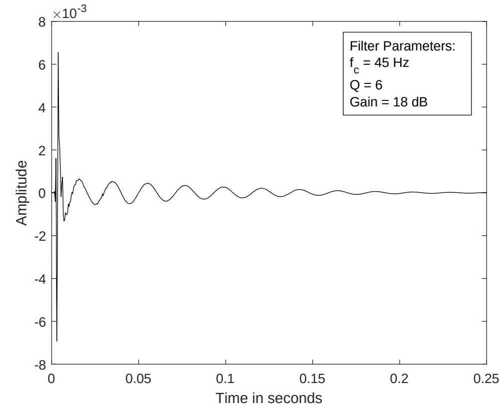
Figure 3.5: Impulse Response, Frequency Response and CSD of Genelec 8020B Monitor Loudspeaker

The drop in magnitude observed in 3.5b is a consequence of the anti aliasing filter approaching the Nyquist frequency, as a consequence for resampling the impulse response.

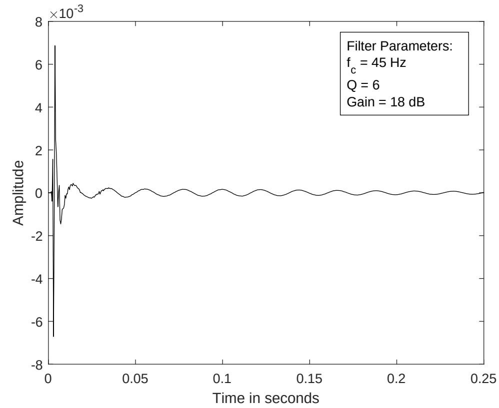
The same filters used earlier with Q values of 6 and 24 are used on the IR, with the same gain of 18 dB. This time, the centre frequency is shifted to 45 Hz instead, since the system's response starts to be flat from 40 Hz, as observed in 3.5b, as well

[3]The procedure and reason for downsampling is explained in Chapter 4

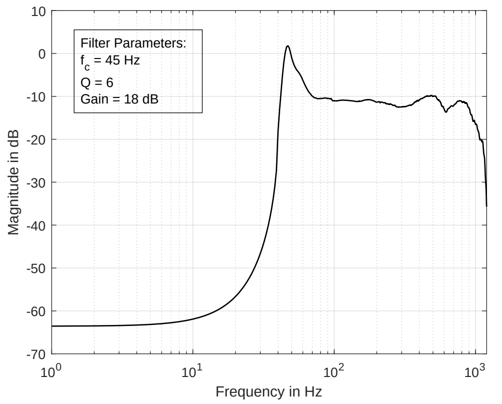
as it is the approximate location of the loudspeaker's mechanical resonance. The following results are obtained.



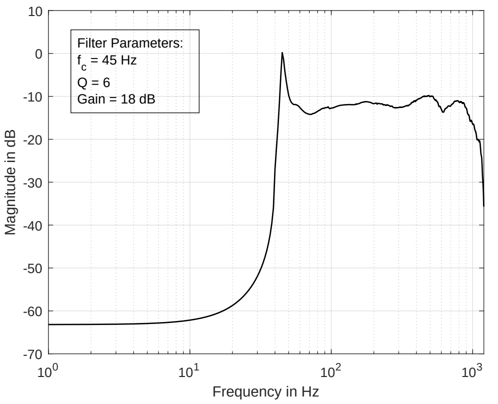
(a) Impulse Response: Q = 6



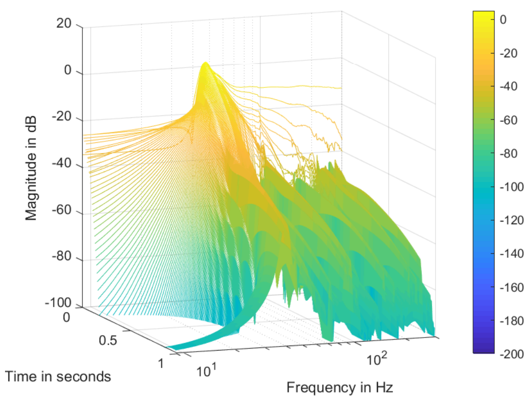
(b) Impulse Response: Q = 24



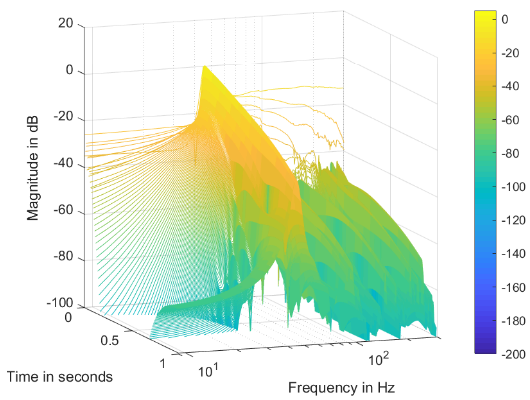
(c) Frequency Response: Q = 6



(d) Frequency Response: Q = 24



(e) CSD: Q = 6



(f) CSD: Q = 24

Figure 3.6: Impulse Response, Frequency Response and CSD of Genelec 8020B Monitor Loudspeaker through filters

As expected, the filters boost the 45 Hz, as well as give a long sustain. The filters also alter the magnitude of other neighbouring frequencies. This is a similar phenomenon observed in room modes, as explained earlier. Frequencies around the centre frequencies will have an altered response due to the strong resonance.

3.4 Perceptual Model to Test Loudspeaker Ringing

What is interesting to observe now is how the unaltered impulse response in figure 3.5a has a very brief sustain, starting at 0.025 milliseconds, for about 30 milliseconds. Its shape is similar to figures A.5a and A.5b, although the filtered responses have been purposely given a strong gain. It could be said that this may be the degree of sustained resonance(i.e., ringing) of the Genelec 8020B. This sustained frequency can be seen in the CSD plot in figure 3.5c prominently. In truth, this is the case for all loudspeakers, since they all will exhibit their own mechanical resonances. Although this may be the case, design aesthetics today have provided solutions to minimize the audibility of these resonances. The Genelec 8020B is a prime example of a well designed loudspeaker. The effect of ringing, although being present, it would be unheard due to the masking effect of audio. Unless an individual has had their ears well trained enough to notice the effect of ringing, this would be unheard, and so the notion of having any importance to the audible effect of loudspeaker ringing has become almost obsolete. This may not always be the case, especially when it comes to the type of loudspeaker, such as car loudspeakers, or PA systems. Monitor loudspeakers such as the Genelec are designed with the thought in mind that it should have a high fidelity and flat response, for the purpose of audio recording. Whereas, car loudspeakers are manufactured to have a generalised specifications. In the sense, it wouldn't pass by as a speaker built for quality and high fidelity, unlike the Genelec.[4]. Most of the reason is attributed to cost effectiveness for a vehicle audio system.

In addition to the quality of a car loudspeaker driver, the placement location of these drivers will also have a crucial influence to the audibility of unwanted resonances, such as cavities behind the driver, uneven structural mountings and additional enclosure mountings in front and back of the driver. Not only these may influence the audibility of the driver, but also may be audible enough to be perceived, thereby affecting the spectral balance of audio.

It is therefore necessary and essential to have a clear conscious and understanding on having a metric on the severity of loudspeaker ringing, as a manufacturing reference. This can aid a manufacturer to understand the cause and effects of ringing, and thereby have a retrospect on this artefact when designing and manufacturing. The main focus on the effect of ringing artefact in the spectral audio will be on its loudness. The methodology employed will correlate the loudness of the artefact with magnitude measurements, utilizing the CSD and Wavelet plots, as well as a

[4]This may not be the case for luxury cars such as Mercedes Benz, as they have Bang & Olufsen (a well known audio company for quality and brilliancy) as their tender for loudspeakers

system identification method involving the Steiglitz - McBride Algorithm [11]. This may need a perceptual model to test the effect of ringing through a series of listening tests, in order to justify the methodology. To undertake this listening test properly, and to have a considerate thought on determining the metric needed for reference, three perceptual models are created, answering three questions:

- **Threshold of Audibility**: What is the minimum level with respect to an audio's level, for a user to be able to perceive the ringing effect
- **Threshold of Equivalence**: What is the level boost required for a non-resonant system to have the same bass energy as a system with resonance
- **Threshold of Flatness**: What is the level cut required for a resonant system to have a flat response

The second and third questions listed above are rather intuitive, as it suggests the energy required to have a perceptive resonance, and the energy required to nullify the audibility of the resonance respectively.

The creation of a loudness metric to the perception of resonance ringing is rather a complex task, as several factors need to be taken into account, such as the dependency on the acoustic environment that can influence the audibility of these resonances [4]. This suggests that several tasks would be required in order to create the metric, wherein each task would have to be looked upon thoroughly and in detail before proceeding to the next task. As a starting point and as the scope of this thesis, it is necessary to have a form of verification to conclude whether the resonance ringing does have any influence to the spectral balance of audio, and to create a methodology to detect and estimate these resonances.

It would be then necessary to have a range of centre frequencies along with its corresponding Q, in order to have an evaluation for several loudspeaker cases, such as a case for sub woofers and a case for mid range drivers. The main frequency of interest is the low frequency resonance, being the mechanical resonance of the loudspeaker drivers as well as the additional resonances from cavities and enclosures in car bodies. It may be expected that the low frequency resonances of drivers would range between 10 - 90 Hz, but this is just a speculative thought, and a more adverse and exact determination of a loudspeaker's resonance is necessary. In order to do so, two methods are followed in which each determines the centre frequency and corresponding Q value. Both methods will be compared, to show their accuracy and reliability. This would require the impulse response measurement of several loudspeakers, in order to show variety. Both the methods as well as the measurement procedure will be explained in detail in Chapter 4.

Upon determining the range of centre frequencies and corresponding Q values, the Listening test model can be implemented, to answer the above mentioned questions needed to evaluate and determine the possibility to create the loudness metric as a reference.

4

Methods

In this chapter, the methodology employed to investigate the perception of resonances caused by cavities and additional mounting structures in a car loudspeaker is discussed. There will be two sections, in which one section focuses on three methodologies to determine the range of centre frequencies and corresponding Q value of several loudspeaker driver resonances. This will give a probable low frequency range in which low frequency resonances are expected. Upon determination on the frequency range with its corresponding Q range, this will serve as input to the second section, in which it will primarily focus on the creation and implementation of the listening test model, in order to understand the levels of perception for resonances.

4.1 Impulse Response Measurement

The first and foremost step into building the listening test model is to determine the range of centre frequencies and its corresponding Q values. This requires the impulse response measurement of several loudspeakers, both Sub-woofer and Mid-Range types. The measurement procedure carried out is based on the works of [5].

The measurement procedure follows the IEC 60268-5:2003 Standard for measurement of loudspeakers in an Anechoic Chamber.

4.1.1 Measurement Equipment

- National Instrument 9234 & 9260 Input/Output Modules
- National Instrument cDAQ 9178 Chassis
- B&K Type 4190 and 4231 Microphone
- B&K Type 2669 Pre Amplifier
- B&K Type 1708 Signal Conditioner
- Genelec 8020B Monitor Loudspeaker and Genelec 7050B Sub Woofer
- Neumann KH80 DSP Monitor Loudspeaker and Neumann KH805 Sub Woofer

The complete detail of the measurement setup, measurement pre conditions, limitations, etc., is described in [5]. This procedure is followed for all above mentioned loudspeakers. The loudspeakers were obtained from the Division of Acoustics, Chalmers University of Technology.

4.1.2 Block Diagram

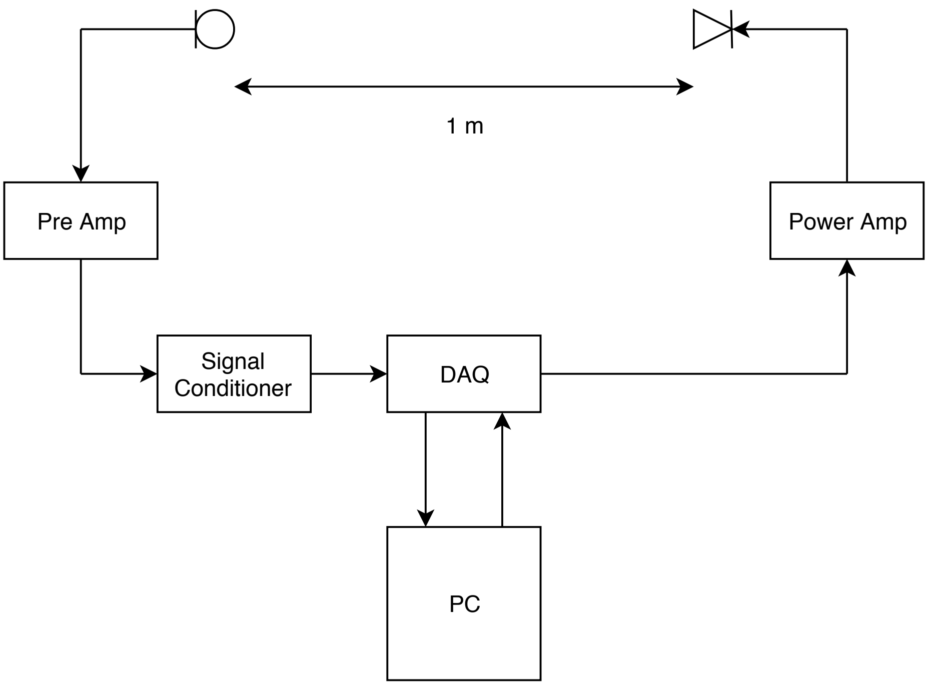


Figure 4.1: Block Diagram of Measurement Setup

4.2 Detection & Estimation of Resonances

The methods used to detect and estimate loudspeaker resonances are based on the works of [1], [2], [3] and [4]. Their works will be split into two methods, in which the results will be compared in the chapter 5.

4.2.1 Method I: Cumulative Specral Decay (CSD) and Continuous Wavelet Transform

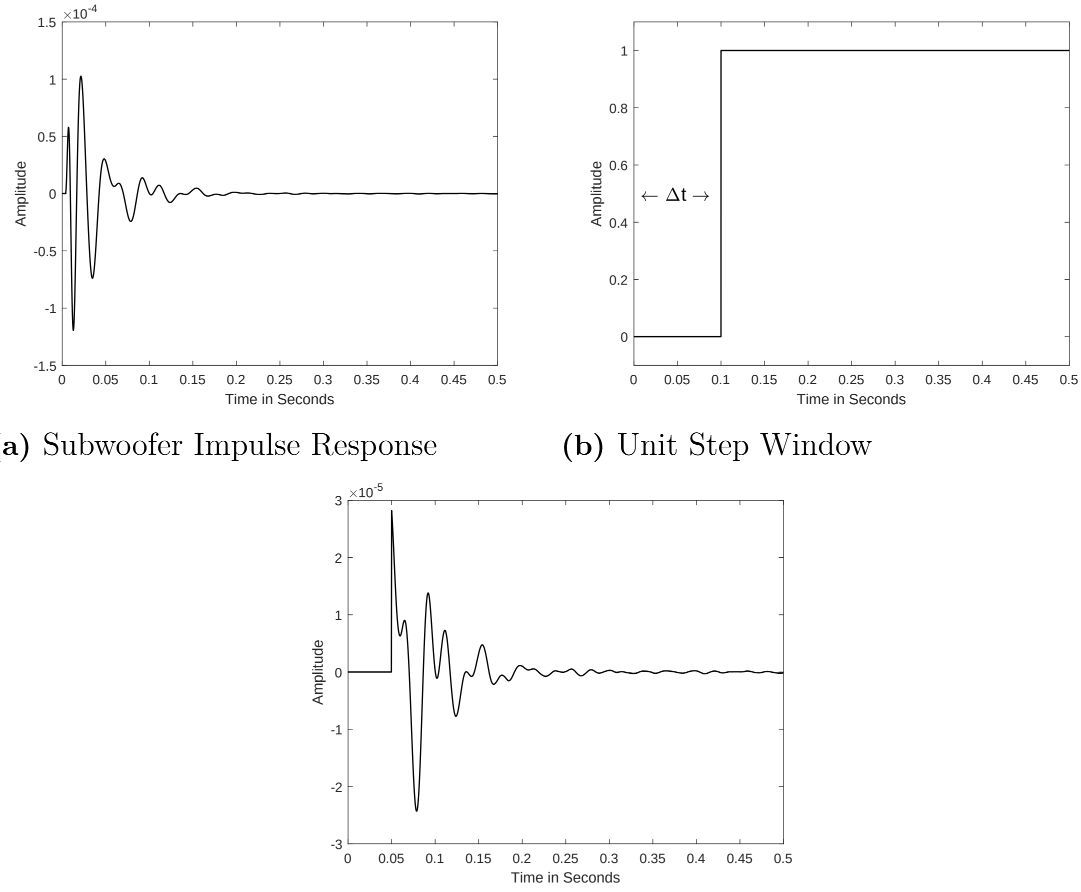
This is a two part section, in which each section describes the method applied to the concept. The process of determining the centre frequency f_c and the quality factor Q are similar for both methods.

4.2.1.1 CSD Method

As described in chapter 2 and 3, the CSD of a signal is a time-frequency analysis that describes the decay profile of frequencies over time. This is very useful when determining the decay profile of a loudspeaker's driver, as it turns out from previous studies in [2],[3],[4] and [8], one can clearly observe the decay profile of a loudspeaker driver's resonance, and other resonances caused by bass reflex ports, harmonic distortions and other artefacts. It will also be useful in observing the decay profile of additional resonances caused by cavities and car body enclosures.

CSD plot essentially represents the magnitude decay over time of each frequency. The CSD of a loudspeaker's driver can be determined from the impulse response measurement described in the previous section.

From the equation described in Chapter 2, the CSD is the FFT of the signal at different time intervals. This can be obtained by multiplying the signal with a series of unit step windows, in which are separated by a time interval Δt. This can be illustrated as follows:



(a) Subwoofer Impulse Response **(b)** Unit Step Window

(c) Sliced Impulse Response

Figure 4.2: Impulse Response,Unit Step Window and Sliced Impulse Response

The signal, being the impulse response of the subwoofer KH805, is multiplied by a unit step window starting at a time Δt, after which the resultant signal shown in figure 4.2c undergoes the FFT process. The resultant response will correspond to the frequency response of the subwoofer, after Δt seconds. So in essence, to compute the CSD, the first slice will correspond to the FFT of the signal with the unit step window at $t = 0$, and the consecutive slices will correspond to multiples of the time interval Δt, i.e., Δt, 2Δt, 3Δt, 4Δt, etc.

Computing the CSD over the whole frequency range is computationally expensive, and would require a large RAM memory in order to process the information efficiently. Moreover, this expense can become larger, depending on the resolution of interest. To ensure a fast computational process, it would be then necessary to resample the signal to a low sampling frequency, since the frequency of interest lies within the low frequency region. Care should be taken, as having a very low sampling frequency may obscure the necessary information. In this case, the signal is resampled to a sampling frequency of 1200Hz, originally being 51.2kHz. Not only the number of samples in a second have drastically reduced, but the computational efficiency has increased extensively. In addition, the resultant frequency response will range from 0 to 600 Hz, which is enough to observe the necessary information. One could reduce the sampling frequency further, so long as the maximum frequency content of interest isn't close to the Nyquist frequency. Moreover, the resampling process implemented in MATLAB makes use of an aliasing filter to filter out the unnecessary decimated points, and a Kaiser window to retain the amplitude of the signal. This would require the user to adjust the resampling settings, in order to preserve the information within the frequency of interest. This will result in the following plots.

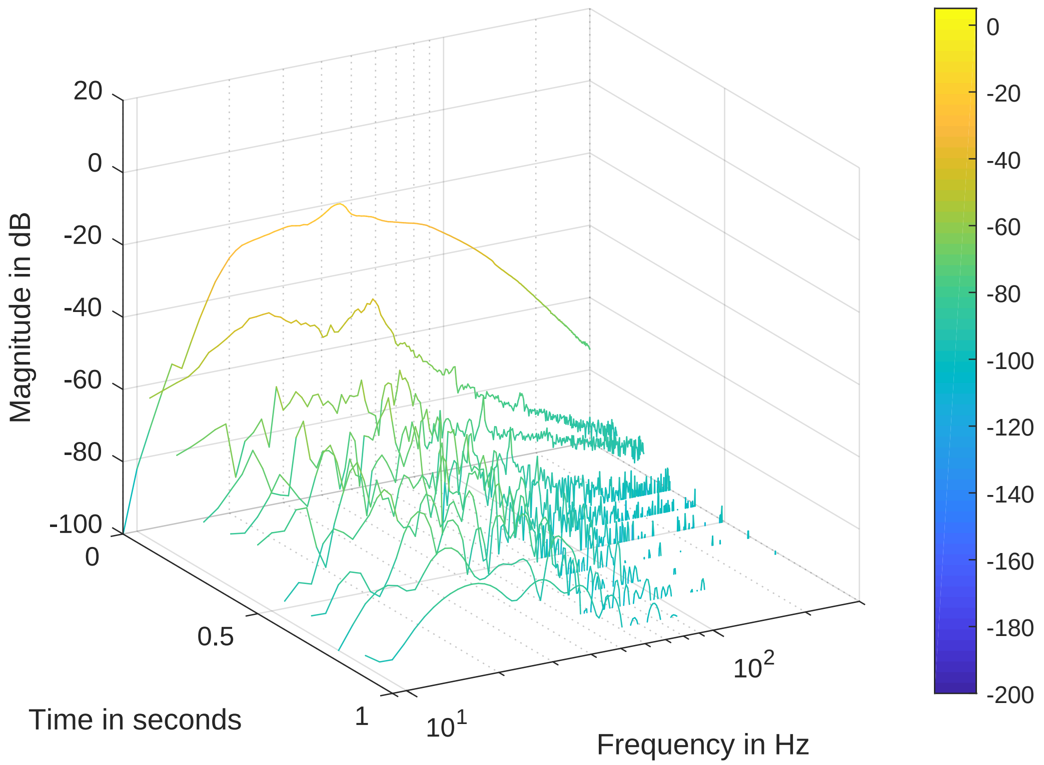


Figure 4.3: CSD of Subwoofer - Δt = 0.1 seconds

These plots are made into waterfall plots, since this kind of plots can visually represent the decay of frequencies over time clearly. The clarity of the content in the plots is dependent on the time interval Δt. For a signal of 1 second, it is important to have a small Δt in order to have enough slices within the length of the signal. For example, if a time interval of 0.1 seconds is chosen, for a sampling frequency of 1200Hz, this time interval will correspond to 120 samples. Hence for a signal of

1200 samples in length, only 10 slices will be obtained. This is illustrated in figure 4.3.

Although one can see the magnitude decay in figure 4.3, the sheer lack of slices makes the information indistinguishable, in terms of events. A small Δt is definitely required, in order to observe the events in the plot. For the same resampled signal, a time interval of $\Delta t = 0.001$ seconds is chosen, which corresponds to 1000 slices. This leads to the following plot:

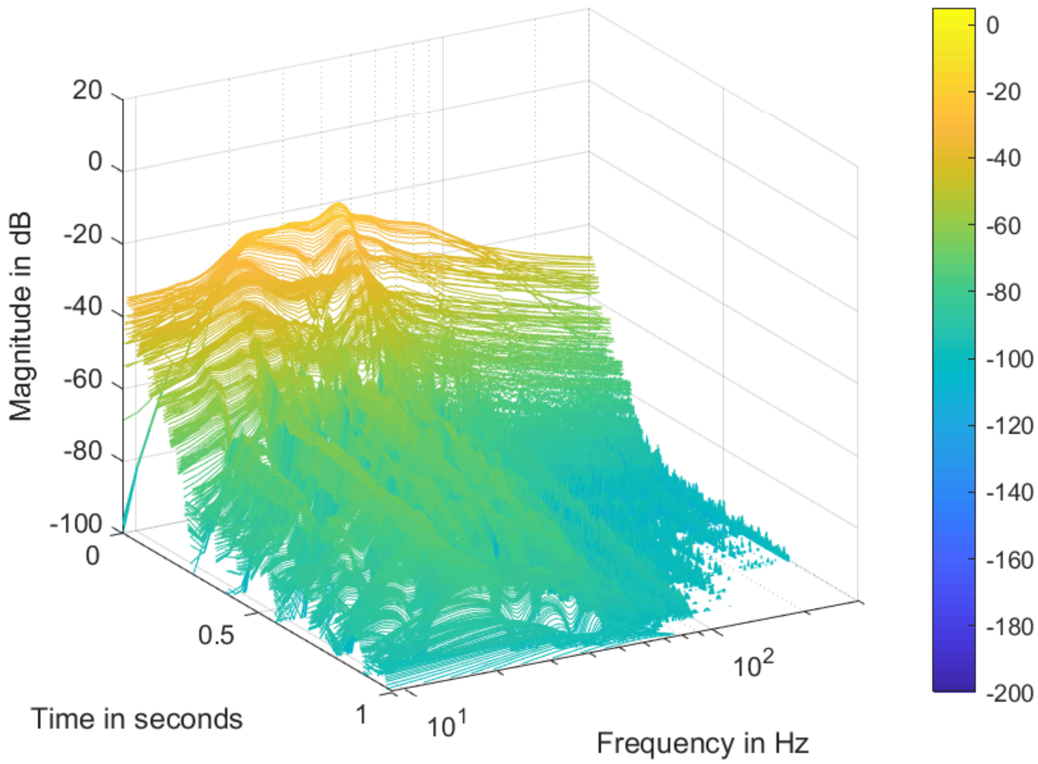


Figure 4.4: CSD of Subwoofer - $\Delta t = 0.001$ seconds

The sheer increase in number of slices show the clarity and resolution of the information shown in the plot.

Although the view and resolution of the plot in figure 4.4 seems to be clear, this can further be improved by modifying the unit step window, a process known as apodization , a concept introduced by John D.Button and Richard H.Small [8]. The authors suggest a method of smoothing the edges and ridges by utilising alternate windows such as a rectangular window, triangular window, Kaiser window, Blackmann window and a Gaussian window. The difference here is that the process of utilising these windows should correspond to the similar use of the Unit Step window, i.e., to have time slices with a time interval Δt. Out of all the windows, the Gaussian window gives the best smoothing outcome. The reason for this modification will be explained in detail in Chapter 5.

Upon applying the modified window, the following plot is obtained:

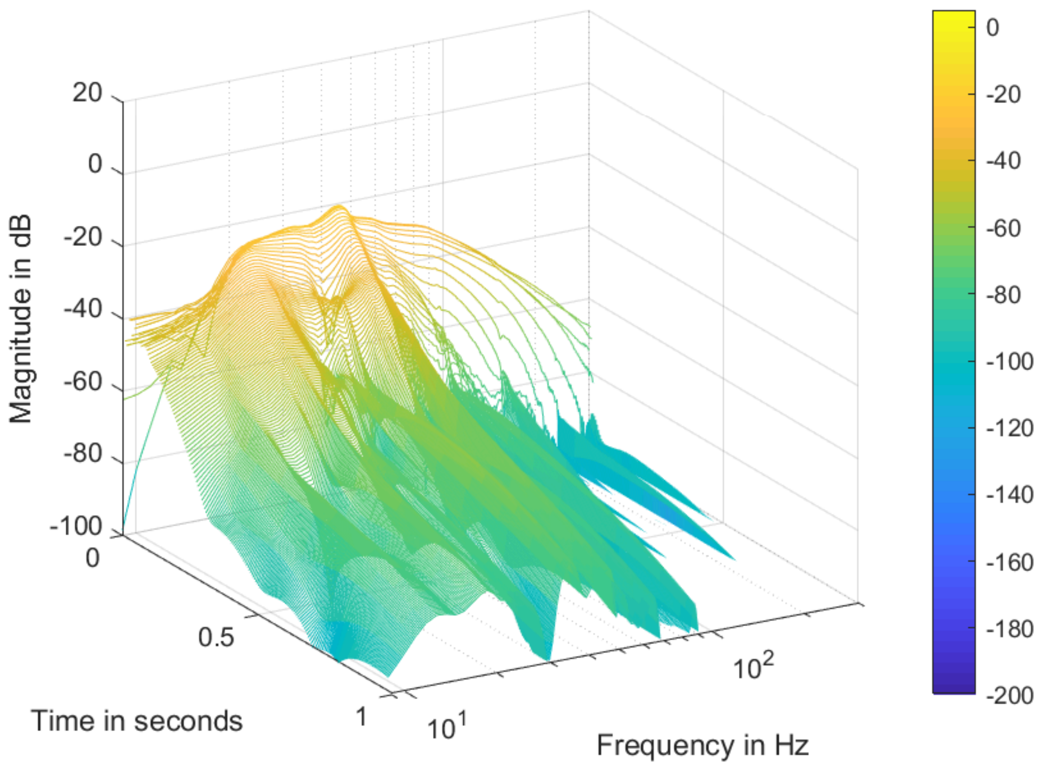


Figure 4.5: CSD of Subwoofer - Δt = 0.005 seconds - Apodized with a Gaussian Window

Clearly, the content in figure 4.5 is much smoother than in figure 4.4, especially when the edges and ridges have been smoothened. In this case, a time interval of 0.005 seconds is chosen. This is because apodization method employed requires the the time interval in samples to be a divisible factor of the total number of samples of the time signal. Previously, the CSD was computed using the heaviside function in MATLAB, which is essentially the unit step function. Any time interval can be used in this case, since the function uses interpolation in order to find the probable location of the time interval within two samples. This process of interpolation is not necessary in implementing the apodization, since the time interval chosen gives a good resolution, as seen in figure 4.5.

The centre frequency of the resonance can be determined at this stage, by calculating the decay times for certain magnitude drops, ranging from a -5 dB drop to -40 dB drop, in steps of 5 dB. This can be achieved by linear regression, in which a line is fit to procure the required drop in level.

The Quality factor Q is calculated, by estimating the frequency bandwidth at a 3 dB drop from the centre frequency. As one might see from the CSD example in figure 4.5, the resonances appear in the plot after a certain time. Considering the frequency response of the entire impulse response, one can't determine the exact 3 db point at the centre frequency, because the magnitude of the adjacent frequencies obscure the resonance. Upon taking a certain slice at a certain time, the shape of the resonance will be clearer, and so the frequency bandwidth can be determined. The Q value can be determined from the following equation:

$$Q = \frac{f_c}{\Delta f} \tag{4.1}$$

$$Q = \frac{f_c}{f_2 - f_1} \tag{4.2}$$

4.2.1.2 Continuous Wavelet Transform (CWT)

The concept of wavelets has been used for many decades, and in recent years, it has found its application in loudspeaker response analysis. The authors in [2],[9] and [10] show a great deal in its application in giving a time-frequency display of loudspeaker responses, as well as a great deal in modal analysis and reverberation time estimations.

The continuous wavelet transform is a form of transform that utilises wavelets of a certain form with a finite length, to condense a time signal into scale coefficients and give a frequency display of events that occur during time [1]. This is proved to be extremely useful, especially in the analysis of seismic data, since one can extract an event individually, and reconstruct the signal with the event alone, excluding other content in the original signal.

The main advantage in CWT analysis is the time-frequency resolution. In general, time-frequency plots, such as in the case of STFT and CSD plots have a fixed resolution. When one increases the time resolution, the frequency resolution reduces and vice versa. In CWT plots however, the resolution is based on the scale length of the wavelet. Wavelets are scaled in length, such that the centre frequency of the wavelet will shift to an equivalent frequency with respect to the time signal. From the equations shown in Chapter 2, the CWT is simply the convolution of the signal with the wavelet. The wavelet will extract the information from the time signal, based on the scale size of the centre frequency. The higher the scale size (i.e. the shorter the length of the wavelet), the higher the frequency range it will extract. This results in a varying time-frequency resolution, wherein low frequencies will have high frequency resolution, and high frequencies will have time resolution. The authors in [2], [9] and [10] propose in using cycle octaves in order to scale the wavelet into powers of 2, as it facilitates in having a relative and fixed bandwidth, and thereby having a fixed scale resolution.

[1] In CWT analysis, the scale coefficients represent the content of the signal at an approximate frequency range

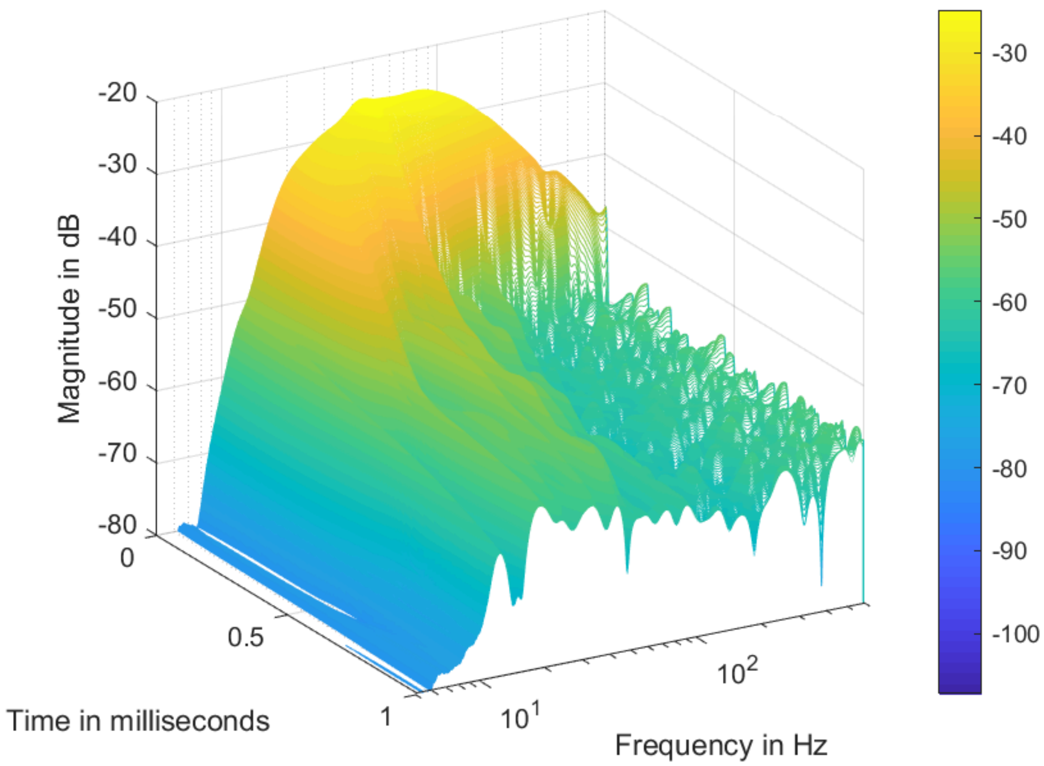


Figure 4.6: CWT of Subwoofer - Voices per Octave = 32

In the case of loudspeaker response analysis, the Morlet Wavelet is used, since its properties satisfies the necessary conditions to extract the information needed in this case [9][10]. The authors in [10] propose to instigate the property of FFT, since convolution can be computationally expensive. In other words, the CWT can alternatively be processed by taking the product of the FFT of the impulse response and the FFT of the wavelet, and taking the IFFT of the final resultant.

The clarity in the content of the plot clearly shows the sustain effect of resonances. The balance in resolution for time and frequency is appropriate enough to visually distinguish the events. At times, it may not be necessary to view the plot in 3D, as transient events can be clearly be distinguishable when viewing in 2D. This can be seen in the figure 4.7.

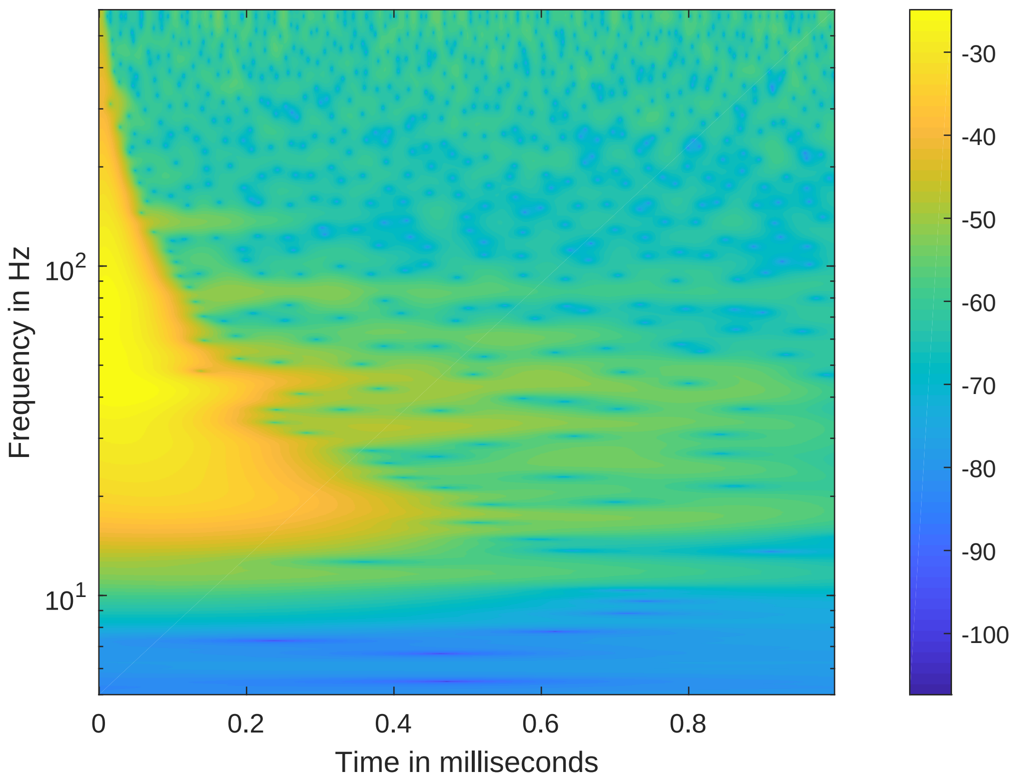


Figure 4.7: CWT of Subwoofer - Voices per Octave = 32

The detection and estimation of the resonances had been unfortunately challenging in this case, and although the CWT plots show the effect of resonances clearly, detecting the resonance through the decay profile method proved to be tricky in this case. The major reason being that the the data points in the frequency slice follows an almost perfect curve. Fitting a line through this data set is impractical. Moreover, since the frequency range is given in terms of scale coefficients, improving the number of frequency bins is redundant.

This has lead to find an alternate approach to detect and estimate the resonances. In this case, a rather robust approach, based on a postulate given in [2] is followed. It mentions that ideal system containing resonances with equal Q and relative equal bandwidth show a true perceptual relevant visualization of these resonances. An example plot obtained from [2] shows the decay pattern of such a system.

The importance in this diagram is the behaviour and visual pattern of the resonances. Having this as the visual reference into identifying a resonance, the estimated CWT is now plotted in 2D, in this case, approximate frequency against magnitude.

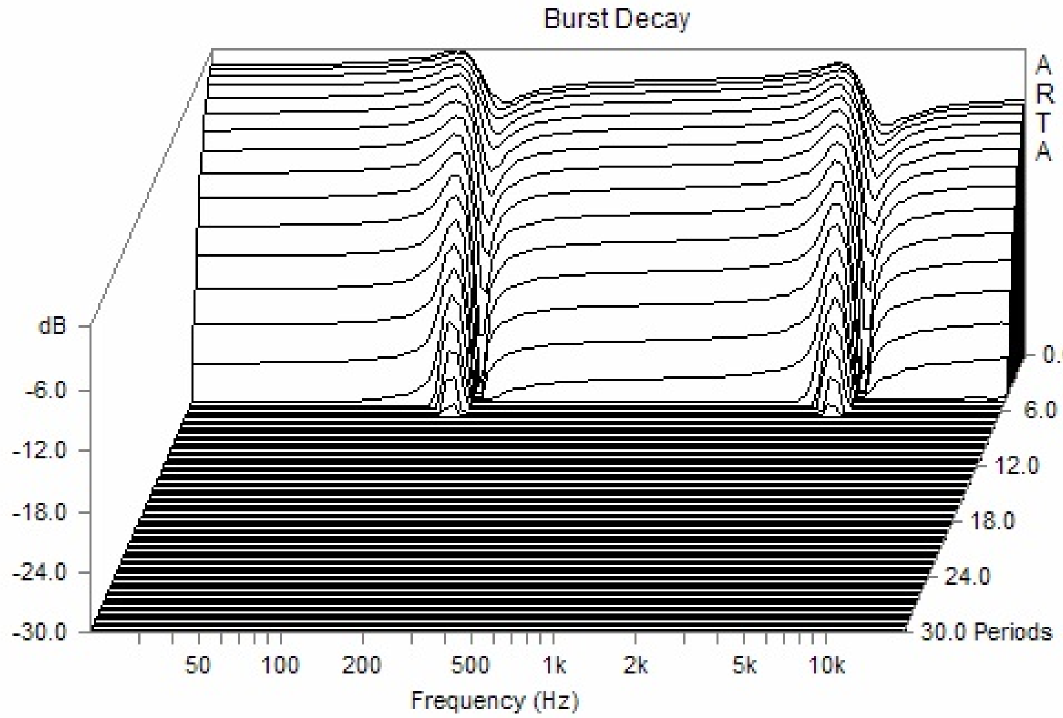
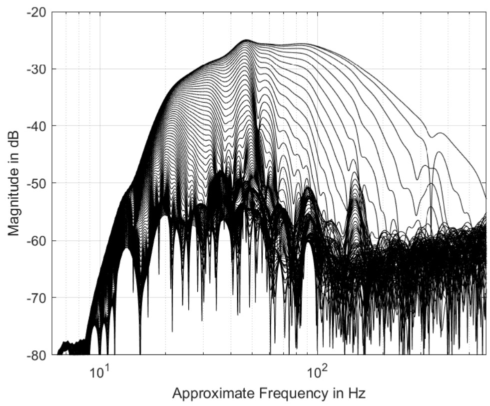
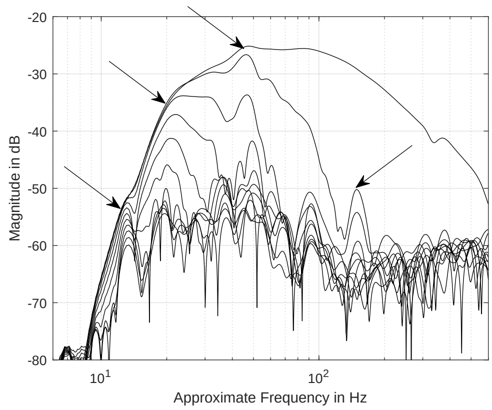


Figure 4.8: Ideal system with two resonances of equal Q - Source : [2]



(a) CWT of subwoofer



(b) CWT of SubWoofer - 12 slices and annotated resonances

Figure 4.9: Visual Identification of Resonances

Figures 4.8 and 4.9a are compared in order to identify the possible resonance locations. Since the high density of slices in figure 4.9a, 12 slices are chosen so as to have a better visualization, as shown in figure 4.9b. The annotated frequencies can be seen in the plot, showing the locations of the resonances. As mentioned already, this method is rather a robust way of identifying the resonances, since it involves a visual interpretation of one, instead of automating the process like in other methods, but since the goal is to verify whether CWT can detect and estimate the resonances, it is fair to follow this method.

Like in the CSD method, a slice after a certain time is taken, and the Q values at the located resonances are estimated. The following table shows the estimated results for all the loudspeakers.

4.2.2 Method II: Pole - Zero Identification Method

This is a methodology, based on the works of [2] and [3], in which an IIR filter can be constructed to have an impulse response equal to the impulse response of any system. In this case, the impulse response of the loudspeaker is used to construct the IIR filter. As explained in [2], the ringing effect of a resonance is caused by the poles of the system. Thereby, if one can determine the poles of the system, the frequency at which these poles correspond to can be identified. This is possible, based on the following derivation:
A resonant system can be characterised by the following transfer function equation

$$H(s) = H'(s) + H_R(s)$$

The resonant part of the transfer function can be described with pairs of complex conjugate poles, $s_p = -\alpha \pm j\beta$. By Partial fractional Expansion of the resonant part of the transfer function, the following equation with a residuum $R = r \pm jv$ is obtained.

$$H_R(s) = \frac{R}{s + \alpha + j\beta} + \frac{R^*}{s + \alpha - j\beta} \tag{4.3}$$

$$H_R(s) = 2r \frac{s + \left(\alpha + \beta \frac{v}{r}\right)}{s^2 + s\frac{\omega_n}{Q} + \omega_n^2} \tag{4.4}$$

where,

$$\omega_n = \left(\alpha^2 + \beta^2\right)^{-1/2} \qquad - \qquad \text{Natural Resonant Frequency}$$

$$Q = \frac{\omega_n}{2\alpha} \qquad - \qquad \text{Q - Factor}$$

From the above equation, upon estimation of the poles, the centre frequency and corresponding quality factor of the resonance can be determined. This requires the estimation of the transfer function coefficients in the S-domain. It is given by the following equation:

$$H(s) = \frac{B(s)}{A(s)} = \frac{\sum_{i=0}^{m} b_i s^i}{\sum_{k=0}^{n} a_k s^k} \tag{4.5}$$

It can be seen that to estimate the transfer function of the resonant component of the system, the knowledge of coefficient values a_n and b_m, as well as the order m and n is needed. Authors in [2] and [3] propose methods to reconstruct the transfer functions through mathematical models and algorithms. The Steiglitz - Mcbride non Linear

Least Squares Estimation method [2] is used in this case. Unfortunately, estimation of poles and zeros in the S - domain is numerically unstable. To counteract this, the estimation of poles and zeros is done in the Z - Domain instead. The corresponding transfer function equation in the Z- Domain is as follows:

$$H(s) = \frac{B(z^{-1})}{A(z^{-1})} = \frac{\sum_{i=0}^{m} b_i z^{-i}}{\sum_{k=0}^{n} a_k z^{-k}}$$

The above equation is can be reduced by partial fraction expansion, in which the constituent residues and poles resemble the transfer function equation given in equation 4.3. The Z-Domain poles can be converted into S-Domain poles by the impulse invariant transformation method, given by the following equation:

$$s_{pk} = \frac{1}{T} \ln z_{pk}$$

From here, the frequency and quality factor for each pole can be determined. From the decay time estimation, the frequencies can be compared with the estimated poles, which will give the corresponding quality factor.
The Steiglitz - Mcbride non LSE method of determining the transfer function is a complex problem, and mathematically solving it can be a tedious task. Fortunately, MATLAB has an in-built function called stmcb() that performs the exact estimation of transfer function coefficients. Unfortunately, the function requires an input order number for the numerator and denominator. The accuracy increases with higher order number. With the given sample rate, $f_s = 51.2kHz$, this will require a huge order number value, and will be computational expensive. As in the case for previous methods, the impulse response is down sampled by decimation, to a sample rate $f_s = 1200Hz$.

This method of determining the resonant frequencies and corresponding Q factor is repeated for different loudspeaker measurements, in which it will give the range of driver resonance frequencies and its corresponding quality factor. The range of centre frequencies and Q factor values are given in the table below:

4.3 Listening Test Model

The listening test model is fabricated to investigate the perception of low frequency resonances by answering three questions: What is the threshold of audibility, How audibly similar does a system without resonances will be if the resonance of another system is added to it, and how flat does the response of a system with a resonance become when a peak filter is used to equalize the resonance.

4.3.1 Simulink Models

In order to provide a proper listening test environment to experiment on the above three questions, three listening test models are fabricated in the Simulink Environment. Simulink is a great platform in providing a good interface environment, in which block models can be fabricated that can be executed on a real time basis. It is widely used in simulations of many processes and systems in the engineering field, and in combination with MATLAB program, it provides a good post processing ability. Given below are the models that will perform three listening tests.

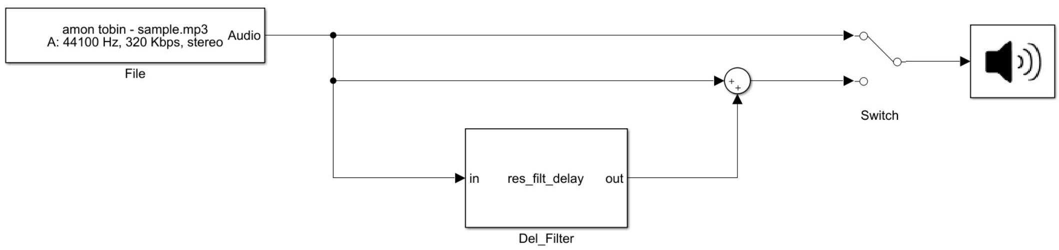


Figure 4.10: Model 1 - Threshold of Audibility Test

The first model consists of an input audio channel, in which two audio signals: a recorded jazz ensemble and a synthetic electronic ensemble, have been chosen. The audio signal is then fed to a resonance filter with parameters estimated from the previous section, and is added back to the original audio signal. The resonances are delayed in time, to emulate the response of delayed resonances with 0 ms amplitude response. The delay times are chosen arbitrarily, but with the conscience of having a real case delayed resonance scenario. The model is integrated with a Graphical User Interface (GUI), that has settings that can change the gain parameter of the resonance, so that the user can adjust the audibility of the resonance, as well as an A/B option to compare the reference audio signal with the Stimuli.

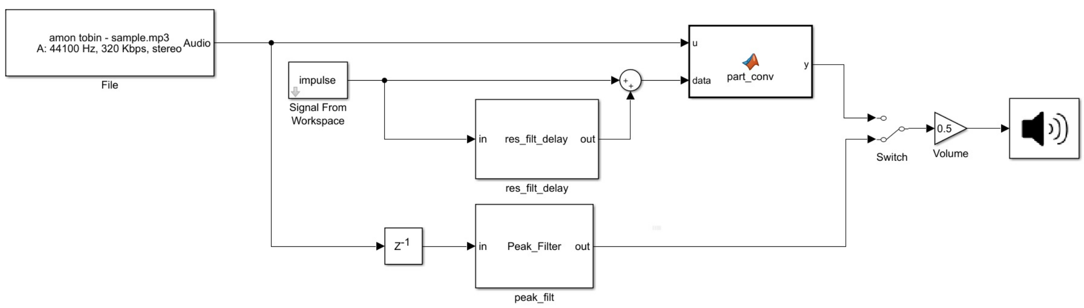


Figure 4.11: Model 2 - Similarity Comparison (Added Resonance)

The second model consists of the same input audio channel. The audio signal is convolved with a Synthetic Impulse Response (SIR), which has the response of the delayed resonances that was used in the first model. This SIR is created by passing an ideal pulse through the same resonance filter used in the first model and is added back to the original pulse. The resultant SIR is convolved with the audio signal by Overlap - Save Fast convolution method, which is fed to a boost peak filter with adjustable gain. The output signal from the peak filter is compared with the original audio signal,ison will suggest the similarity in audible resonance between the two. A GUI will be used for the used to adjust the gain of the peak filter, to suggest the amount of similarity between the two signals.

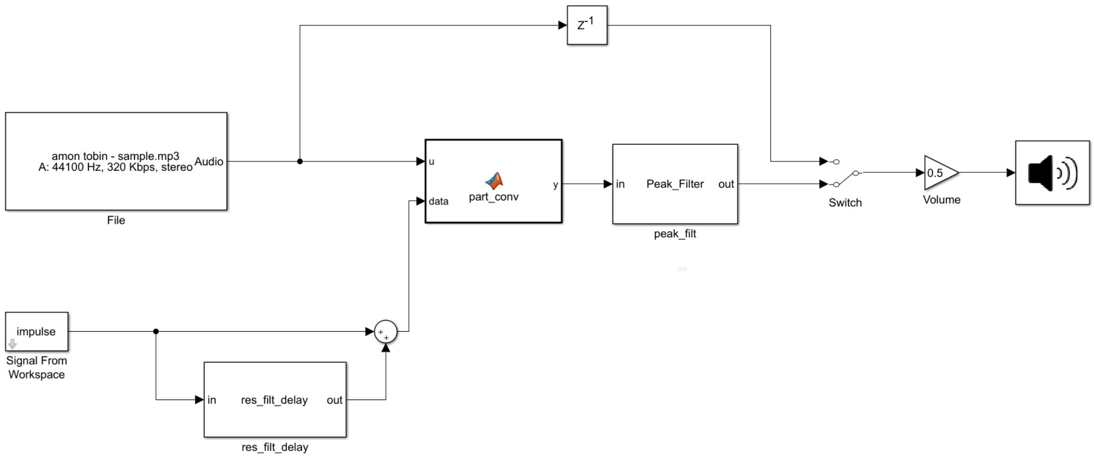


Figure 4.12: Model 3 - Similarity Comparison (Removed Resonance)

The third model is similar to the second model, with the difference that the convolved signal is passed through a Cut peak filter, and on the same GUI, the user will remove the effect of the resonance by changing the gain of the peak filter, and is compared with the original audio signal.

The first model performs the threshold of audibility, and is the first listening test. This will be undertaken separately to the other two tests, as the results of this test will determine the level at which the chosen resonances are audible for the given audio signal. With this information, the initial conditions for the resonant filter will be set, such that it would be clearly audible for the second and third tests.

4.3.2 MATLAB Graphical User Interface (GUI)

The listening test models are operated through MATLAB's Graphical User Interface (GUI), an intuitive feature in MATLAB that allows a user to create an interactive interface with functions that can be customised for any purpose. As this listening test needs to be operated in real time, a control pad to engage Simulink should be created to start and run Simulink in the background. This can be done easily with the GUI.

Two separate listening tests are made. The first one will test the Threshold of

Audibility to find the minimum level at which a resonance is perceived. The second one will test both the Threshold of Equivalence and the Threshold of Flatness, to find the level required for a peak filter to match the level of the resonance and to remove the effect of the resonance respectively.

4.3.2.1 Control Panel for Listening Test I

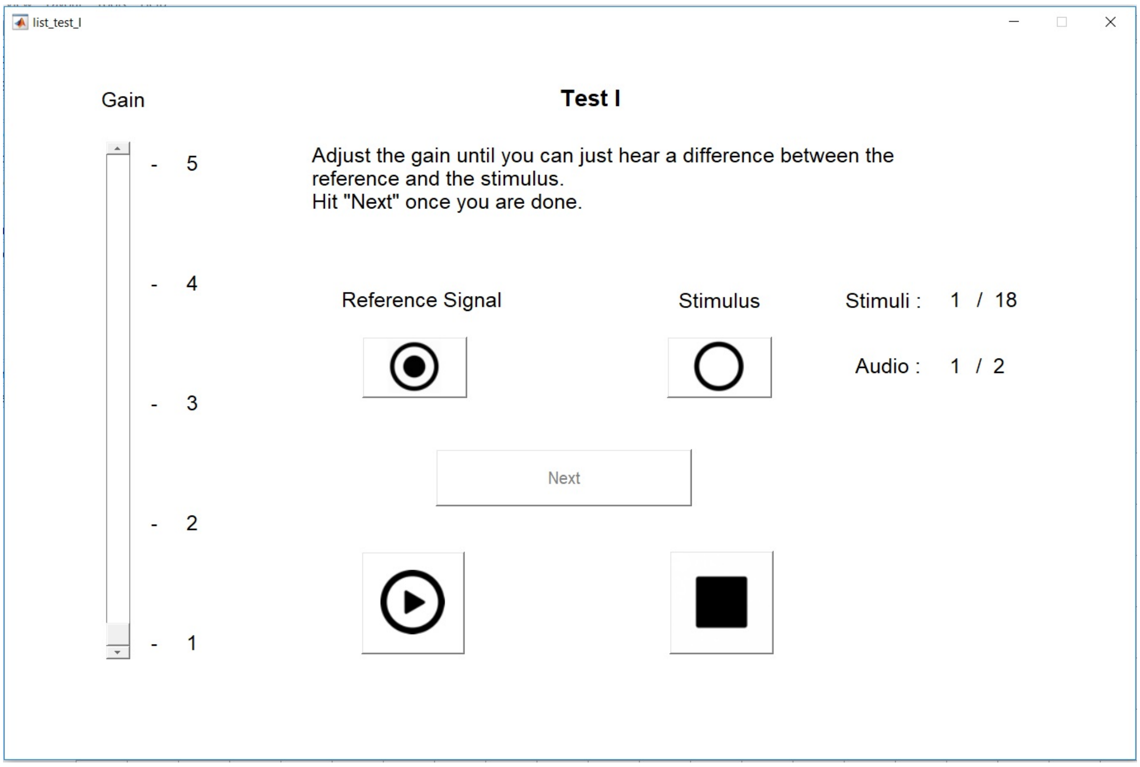


Figure 4.13: Listening Test I - GUI Control

The control panel for the first Listening Test is shown in figure 4.13. It consists of a Gain slider, a reference-to-stimulus toggle switch, playback buttons and a Next Button. This control panel is used for testing and controlling the settings of Model 1, shown in figure 4.10. The gain slider adjusts the gain level of the resonance filter in 4.10, and the reference-to-stimulus switch controls the switch between the unfiltered audio track and the filtered audio track. The Next button saves the gain level value and loads a new stimulus setting to the resonance filter. The markings on the side of the gain level slider are ambiguous to the actual gain level of the resonance filter, as this test would require the level to be less transparent to a user, as part of the experiment.

Unlike conventional A/B comparison listening tests, this listening test runs on real-time. This is an essential part for all tests, since the threshold level is subjected to variations with respect to a subject's hearing sensitiveness. The users will be able to adjust the level of the filter, whilst listening to the change in auditory perception of the test audio. This will greatly facilitate a subject's accuracy in determining their threshold.

4.3.2.2 Control Panel for Listening Test II

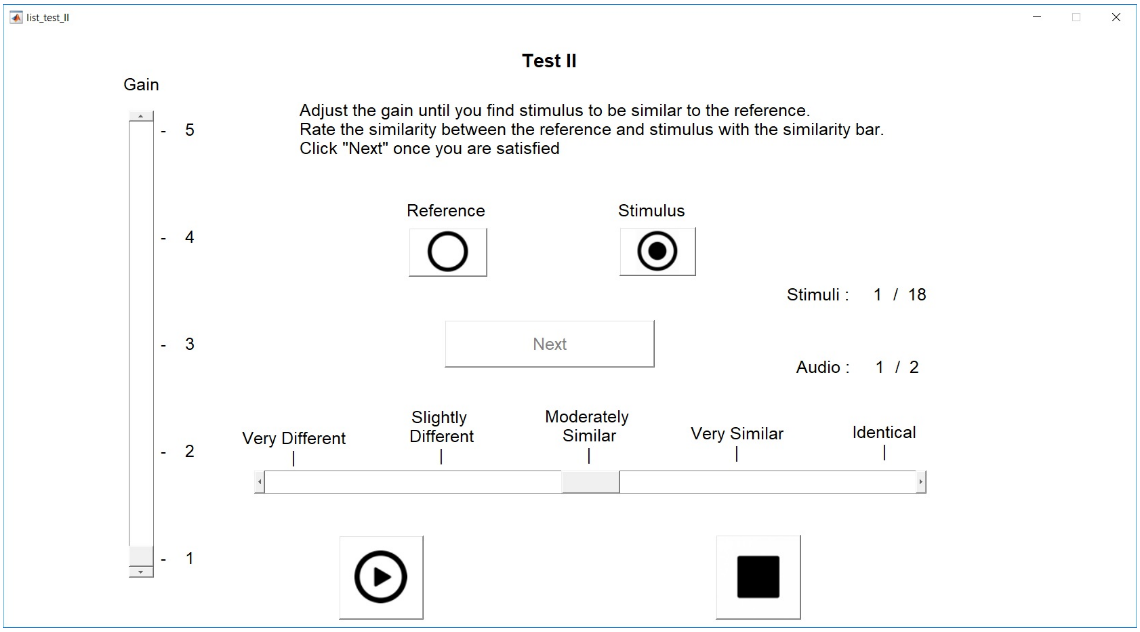


Figure 4.14: Listening Test II - GUI Control

The control panel for the second Listening Test is shown in figure 4.14. This panel has the same controls as shown in 4.13, with the addition of a similarity slider, which is used to rate the similarity between the reference and stimulus audio. This panel is used in controlling the second and third listening models. Moreover, the gain slider in this control panel will be used to adjust the level of the peak filter, instead of the resonance filter. Similar to the previous test, the gain level markings are kept ambiguous, in order to keep opaqueness to the actual level set on the filter, as well as the test is run on real time.

4.3.3 Listening Test Experiment

In this experiment, two separate listening tests are conducted. The first listening test will implement the first listening model, in which it will determine the threshold of audibility. The second listening test will implement the second and third listening models, upon which the threshold of equivalence and threshold of flatness are determined. As mentioned earlier, it is important to conduct the first listening test separately, as the results of this experiment will provide the necessary information to set the minimum gain levels for the resonance filter in the second and third listening model.
For both listening tests, a fixated set of centre frequencies and quality factors are chosen, which are based on the results of the estimated resonance characteristics. For each frequency, three delay settings are chosen, so as to test and emulate the influence of delayed resonances. Two audio signals, one being a jazz ensemble and the other being an electronic ensemble, are chosen for this experiment. This leads to a total of 18 audio samples. To have better variance, the order of the samples

are randomised. Moreover, in order to have surety that the subjects are choosing the right levels without alteration, the audio samples are repeated once, which increases the total number of audio samples to 36. Furthermore, the gain levels of the resonance filter vary randomly with change in audio sample, ensuring that the test isn't preconceived.

In the first listening test, the reference signal will be the raw audio signal and the stimulus signal will be the audio signal with the delayed resonance. The test commences when the playback button is pressed. While the audio signal runs, the subject can switch between the reference channel and the stimulus channel, without the need to stop the playback. The subject can also adjust the gain level of the resonance filter while the audio is running. The task for this test involves the subject to adjust the gain level of the resonance filter, up to a level in which the resonance in the stimulus channel is **minimally heard**. Upon determining the adequate level, the subject can press the next button, after which the chosen gain is recorded and the next resonance setting is loaded, all of which is implemented while the audio signal is running. The task is implemented for all the stimulus settings.

In the second listening test, there are two separate setting involved, depending on the listening model. For the second listening model, the reference signal is the audio signal with the delayed resonance and the stimulus signal will be the audio signal with the boost peak filter. In the case of the third listening test model, the reference signal is the raw audio signal and the stimulus signal is the audio signal with the delayed resonance and the cut peak filter. Like the first listening model, the test can be implemented without stopping the audio playback, and also the subject can switch between the reference channel and stimulus channel without stopping the audio playback. The gain slider will alter the gain level of the peak filter while the gain of the resonance filter is fixated. The task for the subject this time, will be adjusting the peak filter gain, depending on the listening model. For the second listening model, the task involves adjusting the peak filter gain, so as to **match the loudness level of the reference signal**. For the third listening model, the task involves adjusting the peak filter gain so as to **remove the effect of the resonance in the stimulus channel**. Once the subject determines the adequate level for both listening models, the subject can press the next button, after which the chosen gain is recorded and the next resonance setting is loaded.

The listening tests were conducted in the listening test room, at the Division of Applied Acoustics, Chalmers University of Technology. The test directly ran from MATLAB, into the Motu Ultralite MK4 Sound interface, that gives an adequate flat and reasonable sound level. In order to have a clear perception of low frequency resonances, using conventional loudspeakers would not suffice, as the effect of the room may influence the experiment. In order to tackle this, the subjects were made to listen through Harman AKG K702 Headphones, which had a really good flat frequency response, especially in the low frequency region.

Figure 4.15: Listening Test Room

(a) Motu Ultralite MK4 Sound interface

(b) Harman AKG K702 Headphones

Figure 4.16: Listening Test Equipment

5

Results

This section is a two-part result for the following:

- Methodology Results
- Listening Test Results

A detailed explanation for each result will be given.

5.1 Methodology Results

All the three methods show a degree of agreement with each other in estimating the centre frequency and the quality factor with a nominal degree of variance. This can be summarized for each individual method in the following sections, in which a brief discussion on the observations of each method is explained in detail, as well as its limitations are given.

Although the methodology was applied for all the loudspeakers, the following plots and observation shown will be based on the Neumann KH805 Subwoofer, since the events that occur are quite similar with each other. The results for each loudspeaker are given in the Appendix **??**

5.1.1 Cummulative Spectral Decay (CSD)

The centre frequency of the loudspeaker's driver resonance was identified through the decay time profile plot, as shown in figure 5.1.

It may be noticed from figure 5.1 that there seem to be an increase in decay times for larger dB drops. This may be a consequence of the linear fit method to estimate the decay times, as fitting a line through a series of data points will vary in slope, depending on the dB drop value. To understand this postulate, a comparison of two frequencies for different drop levels is given.

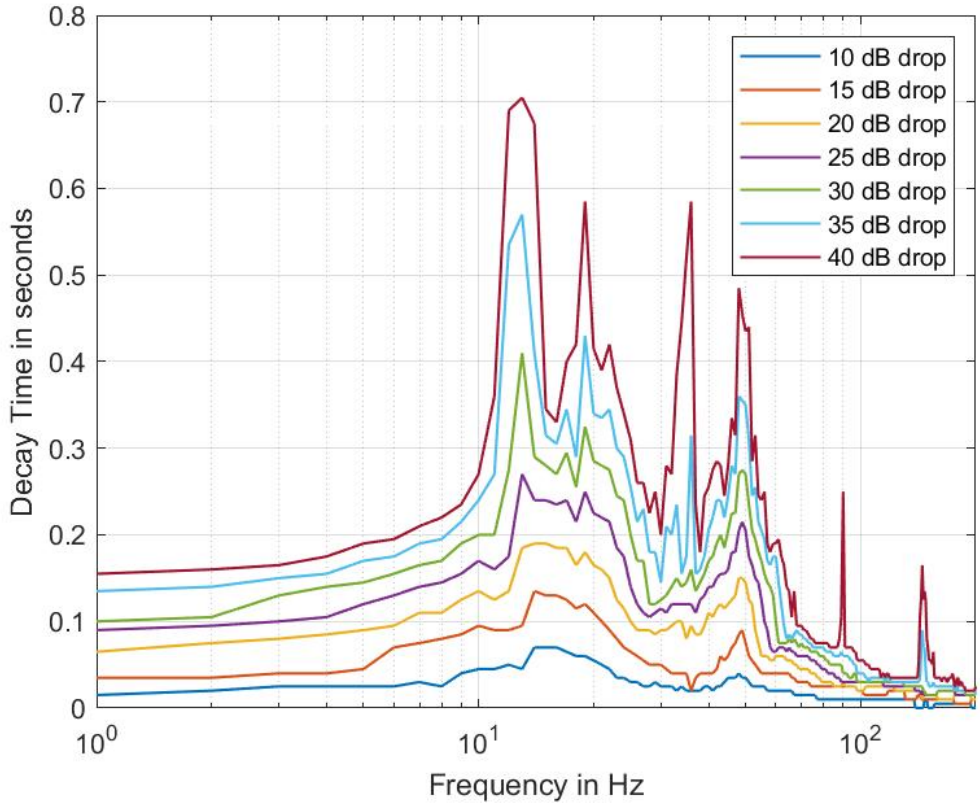
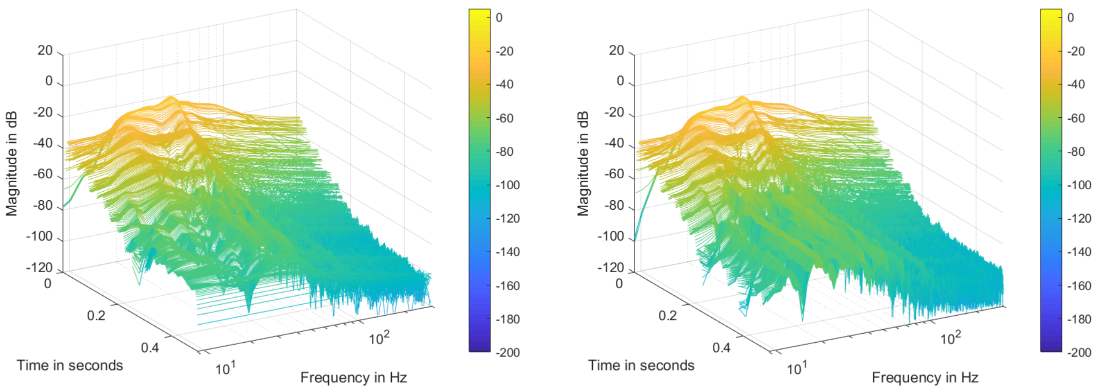


Figure 5.1: Decay time Profile for different drop levels

As the decay pattern differs for each frequency, fitting a line within the set of data points for different dB drops will have different slopes, and for larger dB drops, the slope will drastically be large. As a result, the estimated decay time, which based on the time difference between two points on the line correspondent to the dB drop will not necessarily correspond to the actual drop in level that is observed in the CSD plot. A possible solution to counter this would be to fit two lines instead of one, each line corresponding to the probable slopes of the data set given. This would be the case, if one wants to observe the decay time for large drops in level, but it is not necessary in this case, since the point of interest is to identify the resonances, whose centre frequency tend to have longer decay times in comparison to other frequencies, as clearly observed in figure 5.1.

The accuracy in estimating the centre frequency has a dependency over three major factors. The first factor, being the unit step window used. As postulated in Chapter 4, in order to have a better resolution on the content and events in the CSD, it would be necessary to apodize the window used to compute the CSD, due to the problem of spectral leakage caused by the unit step window [8]. When computing the FFT of a signal, the mathematical function assumes the signal to be periodic over the whole signal length. This means that the moment the signal is altered, the periodicity changes, which will affect the spectral content. Moreover in this case, the abrupt change in signal content caused by the unit step window introduces artefacts into the result, affecting the spectral content. This is clearly seen in figure 1a, wherein the magnitude of certain slices seem to suddenly have larger magnitudes than its predecessor slice, especially in the low frequency region. This is a clear consequence of spectral leakage. Apodization helps to reduce the effect of spectral leakage, by smoothing the abrupt change caused by the unit step window.

The second factor is the length of the time signal. The number of samples within one second can have an effect on the resolution of the events that occur. To show case on this factor, the following plots will be a CSD comparison without apodization:



(a) IR time length = 0.5 seconds

(b) IR time lenth = 1 second

Figure 5.2: CSD Comparison of IR with varying time lengths

Figure 5.2 shows a comparison of the CSD of the subwoofer, with varying impulse response lengths. Clearly between the 0.4-0.5 second time range, an event occurs, that cannot be seen in figure 5.2a. It is crucial to have a long impulse response signal in order to observe the events that can occur, even though the impulse response length of loudspeakers are supposed to be short.

The third factor is the sampling frequency. Although it is beneficial in terms of computational efficiency to downsample the signal to low sampling frequencies, this will in turn will affect the number of frequency bins. This is evident, when observing the decay time plots in figure 5.1, where there are regions of abrupt changes in time length within consecutive frequencies. In order to reduce this abrupt change, higher number of frequency bins are required. This would in turn increase the frequency resolution.

There were a number of challenges and issues that rose in the quality factor estimation. Initially, due the spectral leakage introduced by the unit step window, the resolution of the CSD plot was unclear, and the method applied to estimate the Q factor had given extraneous values that were wrong. This resulted in the application of the apodization technique to improve the resolution, and had given more reliable results.
The following table gives the estimation of centre frequencies their corresponding quality factor for all the given loudspeakers.

Table 5.1: Estimated Centre Frequency and Quality Factor for Loudspeakers using CSD

Loudspeaker	Centre Frequency F_c Hz	Quality Factor Q
Genelec 8020B	44	6.1
	98	10
Genelec 7050B	28	3.1
	46	6.2
Neumann KH805	20	2.9
	46	6.1

5.1.2 Continuous Wavelet Transform (CWT)

The detection and estimation of the resonances had been unfortunately challenging in this case, and although the CWT plots show the effect of resonances clearly, detecting the resonance through the decay profile method proved to be tricky in this case. The major reason being that the the data points in the frequency slice follows an almost perfect curve. Fitting a line through this data set is impractical. Moreover, since the frequency range is given in terms of scale coefficients, improving the number of frequency bins is redundant.
This has lead to find an alternate approach to detect and estimate the resonances. In this case, a rather robust approach, based on a postulate given in [2] is followed. It mentions that ideal system containing resonances with equal Q and relative equal bandwidth show a true perceptual relevant visualization of these resonances. An example plot obtained from [2] shows the decay pattern of such a system.
The importance in this diagram is the behaviour and visual pattern of the resonances. Having this as the visual reference into identifying a resonance, the estimated CWT is now plotted in 2D, in this case, approximate frequency against magnitude.
Like in the CSD method, a slice after a certain time is taken, and the Q values at the located resonances are estimated. The following table shows the estimated results for all the loudspeakers.

Table 5.2: Estimated Centre Frequency and Quality Factor for Loudspeakers using CWT

Loudspeaker	Centre Frequency F_c Hz	Quality Factor Q
Genelec 8020B	41	5.5
	98	10
Genelec 7050B	15	2.7
	46	4.7
	90	9
Neumann KH805	13	5.6
	21	2.6
	49	6.3
	90	9.5

5.1.3 System Identification Method

As per the procedure shown in Chapter 4, the transfer function of the loudspeaker was identified, which lead to the following reconstructed impulse response.

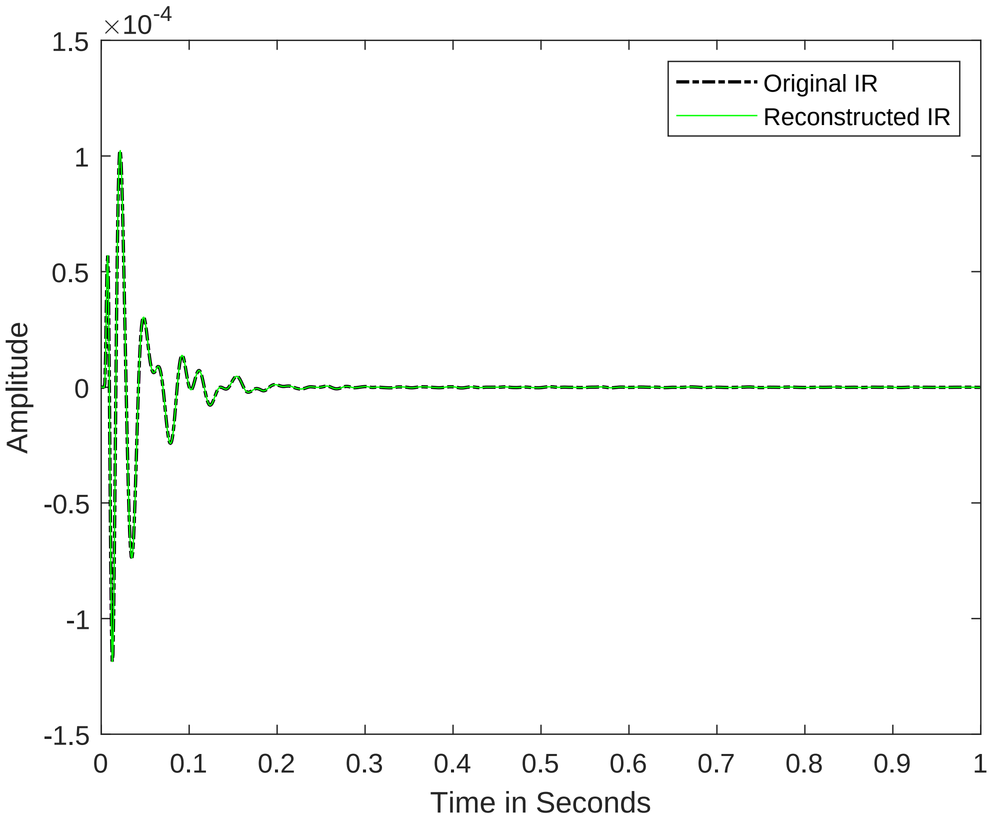


Figure 5.3: Comparison of Original impulse response and reconstructed impulse response

As mentioned in Chapter 4, the Steiglitz-Mcbride Algorithm was implemented by using the in-built stmcb() function in MATLAB, which required to give an input numerator and denominator order. Since the impulse response signal used has been downsampled to 1200 Hz, the maximum order number that gave the best reconstruction signal was 1200. The numerator order was chosen to be 1170 and the denominator order was chosen to be 1200. The reason why the order numbers weren't the same is because this would have created an FIR filter response with no poles. It is necessary to make sure that the chosen numerator order is lesser than the denominator order. Figure 5.3 shows the comparison between the original impulse response and the reconstructed impulse response. The algorithm had almost perfectly reconstructed the impulse response, although it seems to show an exact match. The slight difference can be seen in the frequency response , as shown in figure 5.4

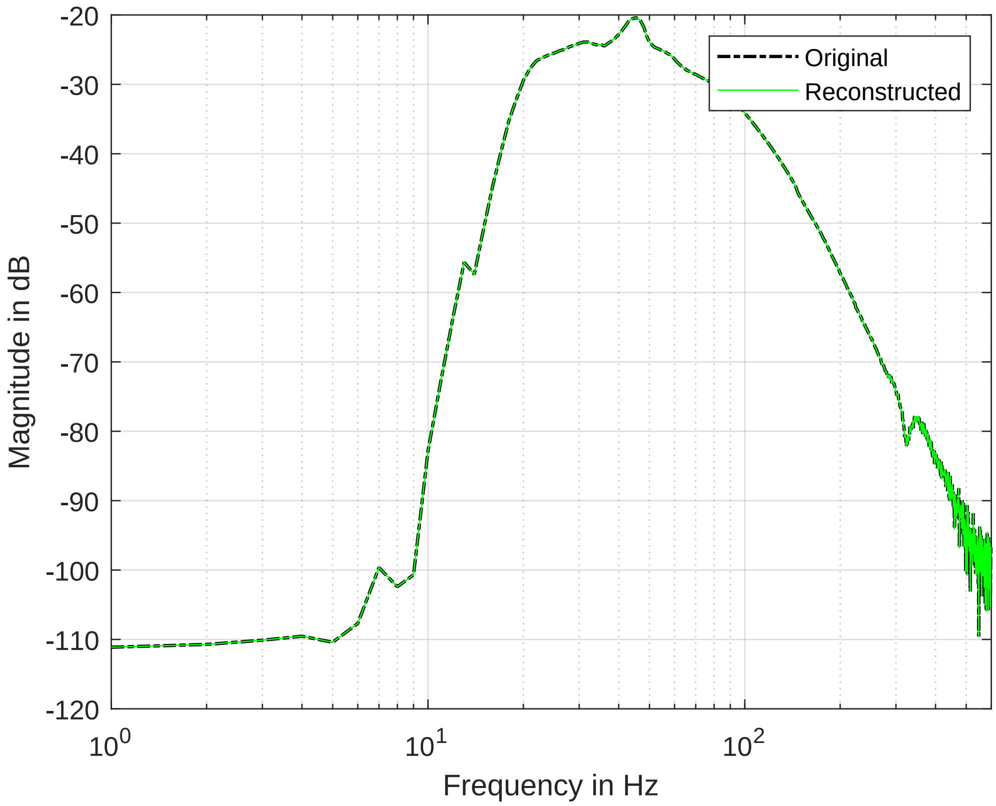


Figure 5.4: Frequency response comparison of Original impulse response and reconstructed impulse response

The minute differences can be seen clearly, although the frequency range of interest shows an exact match, which is the required need in this case. Applying the equations given in Chapter 4, the centre frequency and corresponding Q value were estimated. At this stage, it is fair to express that a problem will arise, especially when there are 1200 poles, giving 1200 frequencies with corresponding Q factors. The obvious problem being, how to segregate the redundant pole frequencies, and identify the true resonance frequency. One solution given in [2], is to use the Single Value Decompostion method to reduce the order number, by decomposition the transfer function into two singular matrices and a diagonal matrix, and estimating the residue of the diagonal matrix. The diagonal matrix gives the true order number, and values a transfer function. Another solution to this is to use the CSD slice at the time chosen in the CSD method. The reconstructed signal created with the estimated transfer function will correspond to the given number of poles and zeros that have been arbitrarily given as input into the function. When observing the estimated Q values, one will see that the values are really high, and none have Q values below 15. This is because the estimated poles are matched to the impulse response given, so as to give the consequent frequency response. This means that if one could reverse the CSD slice chosen in the previous section into a time signal, this time signal will correspond to the impulse response of the apparent frequency response of the chosen CSD slice. Since the resonances are much evident after a certain decay time, it would then be possible to estimate poles to match the impulse response given, that give the required response.

The SVD method was attempted, but due to unexplained circumstances, it was rather difficult to implement the method. The CSD slice method however, proved to be easier. The IFFT of the chosen CSD slice was taken, to give a time signal, that corresponds to the impulse response of the slice. This was used as input signal into the function, and the procedure shown in Chapter 4 was followed. This lead to the following plots:

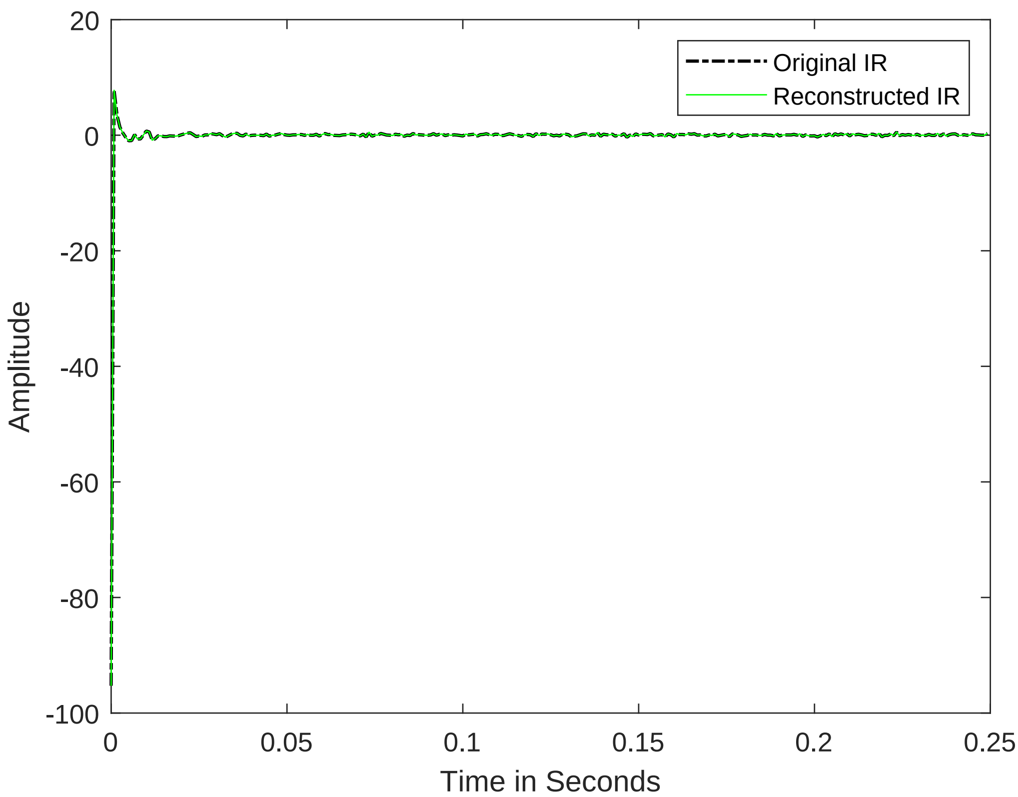


Figure 5.5: Comparison of Modified impulse response and reconstructed impulse response

The strange artefact seen in the beginning that has such a low amplitude, may be a consequence of apodization, since the CSD slice was smoothened. Regardless, the method almost accurately recreated the impulse response as shown in figure 5.5. The frequency response is as follows:

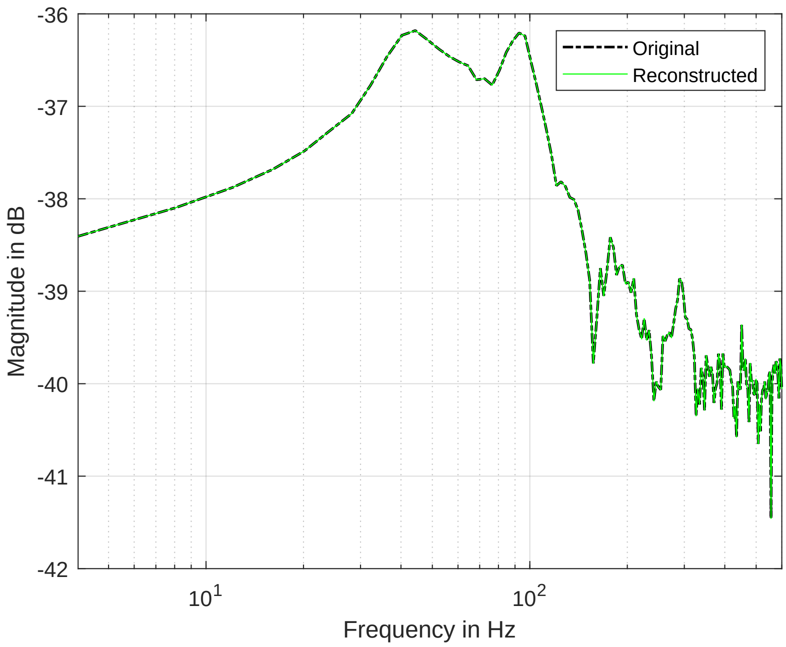


Figure 5.6: Frequency Response Comparison of Modified impulse response and reconstructed impulse response

As expected the estimated poles now match the impulse response and give an almost accurate frequency response. The frequencies and corresponding Q values can now be estimated from the poles, and by matching the frequency at the peaks on the plot in figure 5.6 with the estimated frequencies, the centre frequency of the resonance and its Q value is obtained. This procedure is followed for all the loudspeakers, and the following table is obtained.

Table 5.3: Estimated Centre Frequency and Quality Factor for Loudspeakers using CSD

Loudspeaker	Centre Frequency F_c Hz	Quality Factor Q
Genelec 8020B	44	5.3
Genelec 7050B	28	3.8
	46	8.1
Neumann KH805	20	3
	46	8.3

5.1.4 Comparison of All Methods

Table 5.4: Comparison of All Methods

Method	Loudspeaker					
	Genelec 8020B		Genelec 7050B		Neumann KH805	
	Centre Frequency	Quality Factor	Centre Frequency	Quality Factor	Centre Frequency	Quality Factor
CSD	44 Hz	6.1	28 Hz	3.1	20	2.9
	98 Hz	10	46 Hz	6.2	46 Hz	6.1
CWT	41 Hz	5.5	15 Hz	2.7	21 Hz	2.6
	98 Hz	10	46 Hz	4.7	49 Hz	6.3
Sys. Ident. Method	-	-	28 Hz	3.8	20 Hz	3
	44 Hz	5.3	46 Hz	8.1	46 Hz	8.3

The common frequencies that have been estimated in all methods are tabulated, to show the trend in their estimations. As it can be seen, in almost all the loudspeakers, the methods were able to pin point the resonant frequencies, and were able to estimate to approximate quality factor.

5.2 Listening Test Results

From the results in the detection and estimation methods, the listening test model is given three frequencies: 35 Hz, 60 Hz and 90 Hz and their corresponding Q value given is 5, 6 and 8 respectively. The listening test was performed for 10 subjects for Test I, and 12 subjects for Test II and III.
The results obtained are categorised based on the three case scenarios as explained in chapter 4. For simplicity, the case scenarios will be termed as:

- Test I - Threshold of Audibility
- Test II - Threshold of Equivalence
- Test III - Threshold of Flatness

The listening results were sorted and made into boxplots, a simple way of representing statistical data on a plot in which a rectangle is drawn to represent the second and third quartiles, usually with a vertical line inside to indicate the median value. The lower and upper quartiles are shown as horizontal lines either side of the rectangle.
At the end of the tests, a verbal query on the experience of the tests were given, in regard to easiness and comparison between the audio samples used.

5.2.1 Test I - Threshold of Audibility

The results for Test-I show that the relative level range between the audio level and the resonance is between 3 - 6 dB, with the exception of Audio 1's result for 35Hz, as shown in figure 5.7.

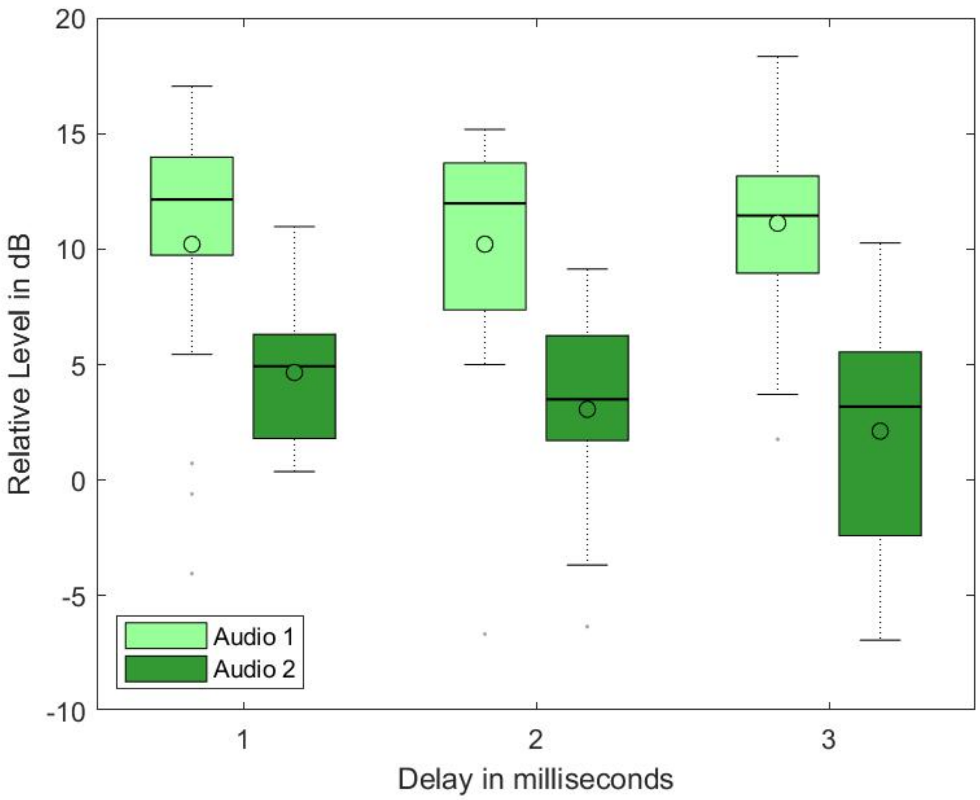


Figure 5.7: Test I Result for 35 Hz

The most probably cause for this wide difference is due to the fact that 35 Hz is rather difficult to perceive if the ambient sound of the audio sample have high reverberance and transients, which happened to be the case for Audio 1. The subjects experienced a form of difficulty in determining the threshold specifically for this audio, due to the extraneous amount of transients and events occurring in the audio.

The difficulty experienced for Audio 1 is also evident for the other frequencies, as shown in figure 5.8 and 5.9, considering the range in variance, which is within 12 dB, and some cases even larger. However, the subjects felt that audio 2 was much easier to determine. This is to be expected, since Audio 2 has no reverberance, and being an electronic ensemble, it is synthetically created.

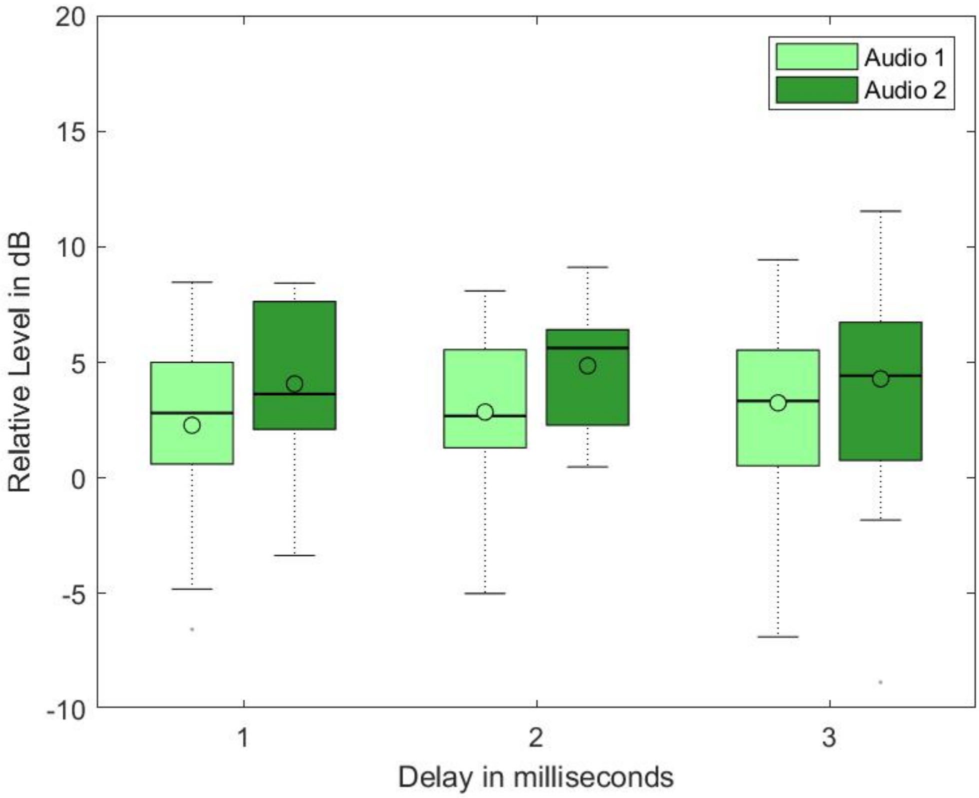


Figure 5.8: Test I Result for 60 Hz

This suggests how the choice of audio can vary the threshold of audibility. The difference in threshold range for different audio obtained here is in par with the results of Floyd and Olive [1], whom have tested subjects with white noise, pink noise, an orchestra and a pop song. The estimated thresholds in [1] vary with audio and frequency, the most noticeable variance being the low frequency. Although this may be the case, there is still a noticeable variance in relative level for all cases. This may be attributed by the number of subjects that have participated, and the hearing capability of the subject. The estimated variable thresholds does not necessary indicate that it corresponds entirely to the type of audio used, although the 25% percentile and 75% percentile have a maximum range of 5 dB at most. This can be regarded as an acceptable threshold value to be used as reference for Test II and Test III.

Considering that the minimum level needed to perceive the 35 Hz frequency in Audio 1 is 10 dB, the resonance level for Test II and Test III were set to 18 dB. This value itself was not enough to perceive the 35 Hz in audio 1, as it will be indicated in the results, but having any higher levels can pose a danger of perceiving loud distortion over other frequencies, since their minimum perceptual level were much lower. It could also damage the AKG500 Headphones, affecting the listening test.

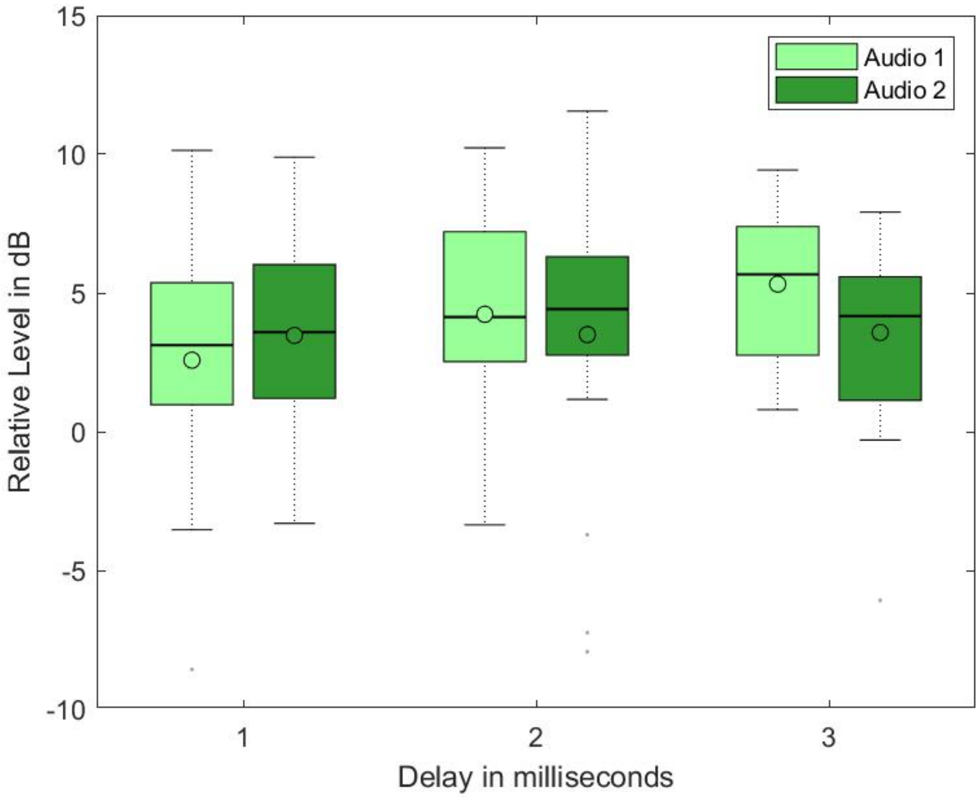


Figure 5.9: Test I Result for 90 Hz

5.2.2 Test II - Threshold of Equivalence

The results of Test II show that the estimated rise in gain level to have a non resonant system to be equivalent to a resonant one is around 19 dB. This is somewhat expected, since the settings of the peak filter have the same centre frequency and quality factor values as the resonance filter. What was was not expected to see is the high variance in perceptual similarity, especially with the 35 Hz in Audio 2, as shown in figure 5.11.

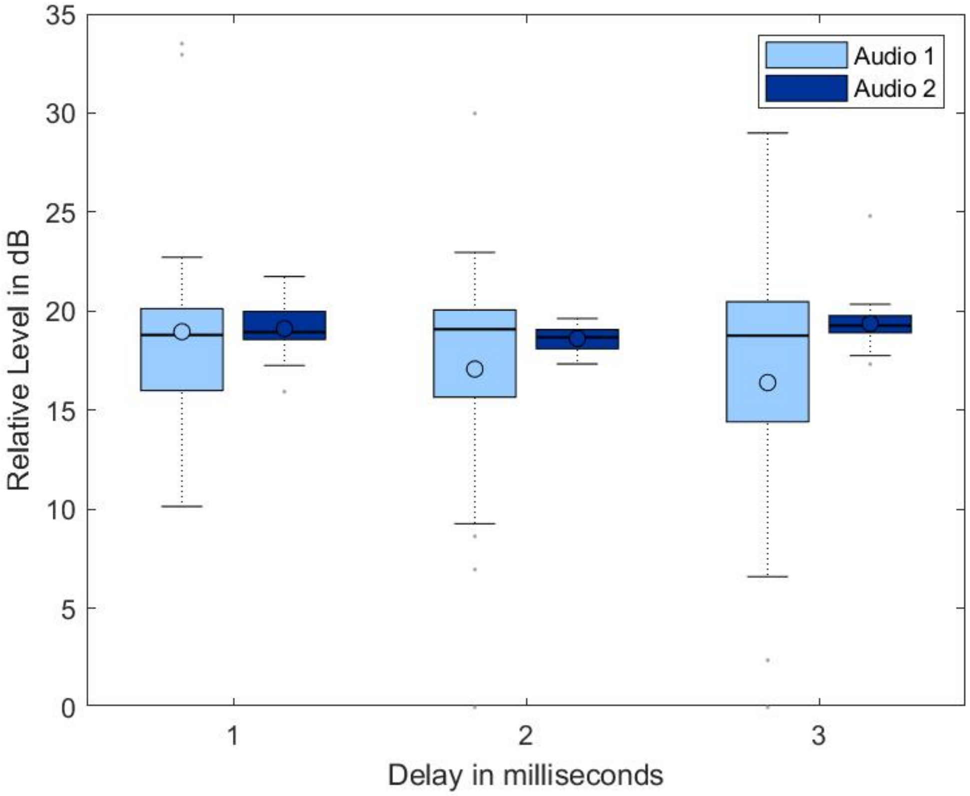


Figure 5.10: Test II Gain Result for 35 Hz

The results in figure 5.10 show the similar postulate in the dependency of audio in perceiving the frequency. The variance in Audio 1 shows otherwise. The variance in Audio 2 is very small, indicating the ease of being able to match the level. In addition, since the perceptual threshold of audio 1 for 35 Hz was estimated to be 10 dB, it was clear that the subjects would again find difficulty in perceiving the 35 Hz resonance in audio 1 at 18 dB. The surprising fact is how the subjects found the 35 Hz in audio 2 to be moderately similar to the resonance, in comparison to audio 1. Initially, the expectation was to have a similarity index to be around 8 to 9, which was the case for every other frequency. Upon receiving this result, a cross check on the perception of the audio 2 for 35 Hz had to be made in order to ensure that this was the actual case. Although as a biased subject, it was surprising to find out that there was a small hint of modulation at this frequency, with respect to audio 2, even though the initial thought was considering that the high variance in the percentiles itself may have been the outcome for having lesser number of subjects.

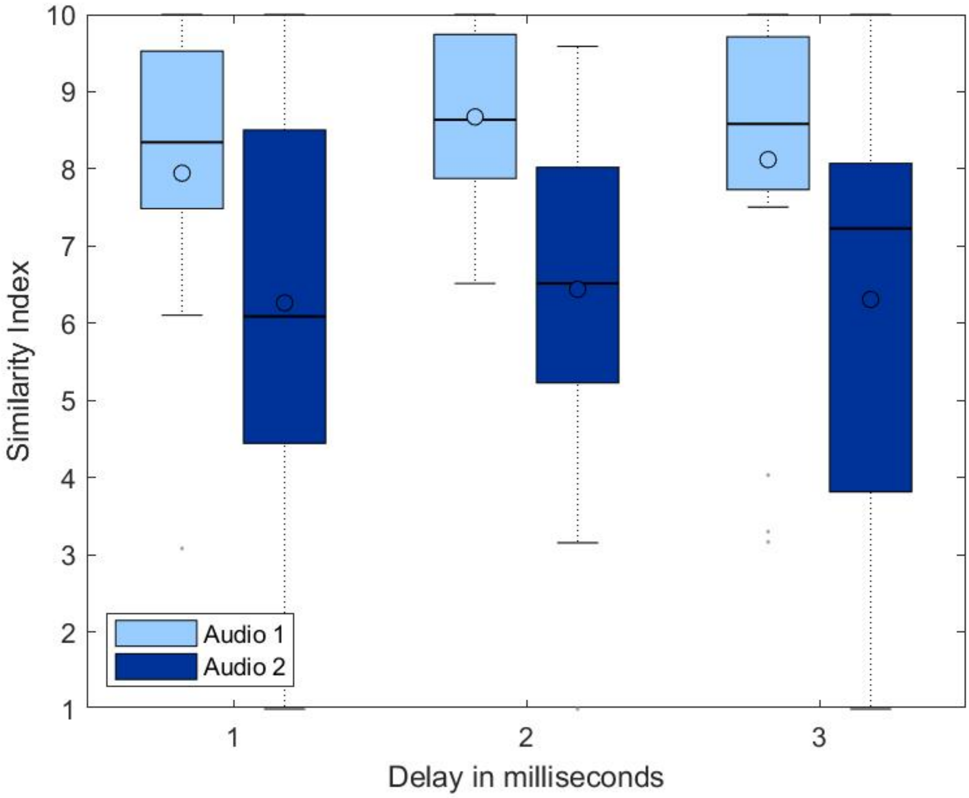


Figure 5.11: Test II Similarity Result for 35 Hz

This may be an indication that the ringing effect of very low frequencies might perceptually modulate the audio. As a matter of opinion, this could be just a special case in regard to synthesized audio, such as Audio 2, since this audio contained really low frequency bass transients. Moreover, this finding cannot be conclusive, as it needs to have its own separate verification in order to have any conclusion out of this.

In the case for 60 Hz, a wide variance is observed in the similarity plots for both audio samples as shown in figure 5.13. This high variance may be the either due to the light number of subjects or due to the sheer volume of the resonance, since the level is high with respect to the threshold. This could be the case, since the similarity indices shown in figure 5.13 are as expected.

In the case for 90 Hz, audio 1 shows to have a higher variance in comparison to audio 2, which is seen to have a nominal variance. This could be regarded as normal, since 90 Hz is regarded to be easily audible than the lower frequencies. This might not be the case in audio 1, as postulated that the recording sound track can mask the frequencies.

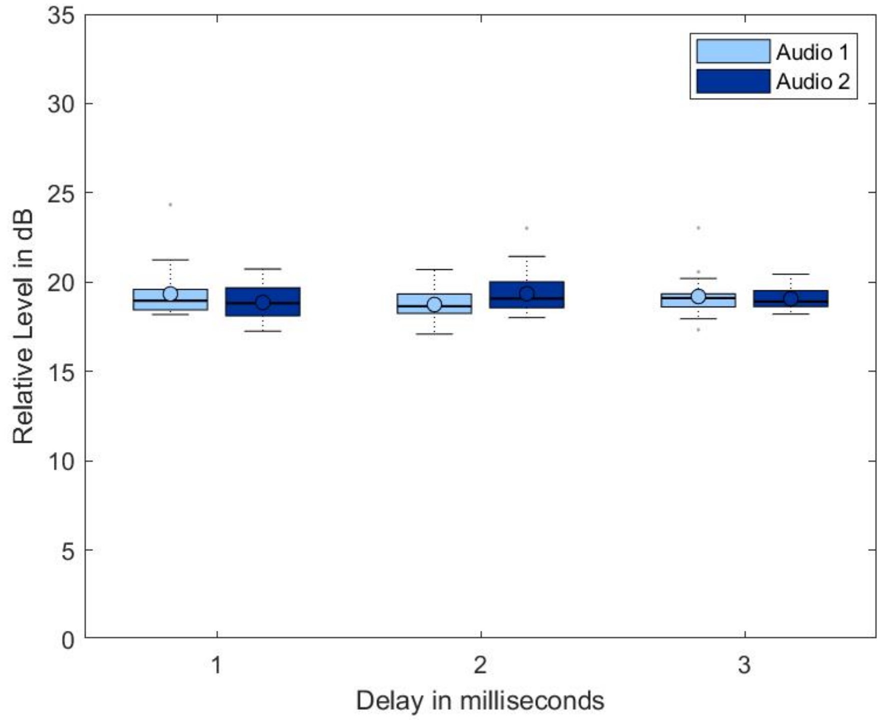


Figure 5.12: Test II Gain Result for 60 Hz

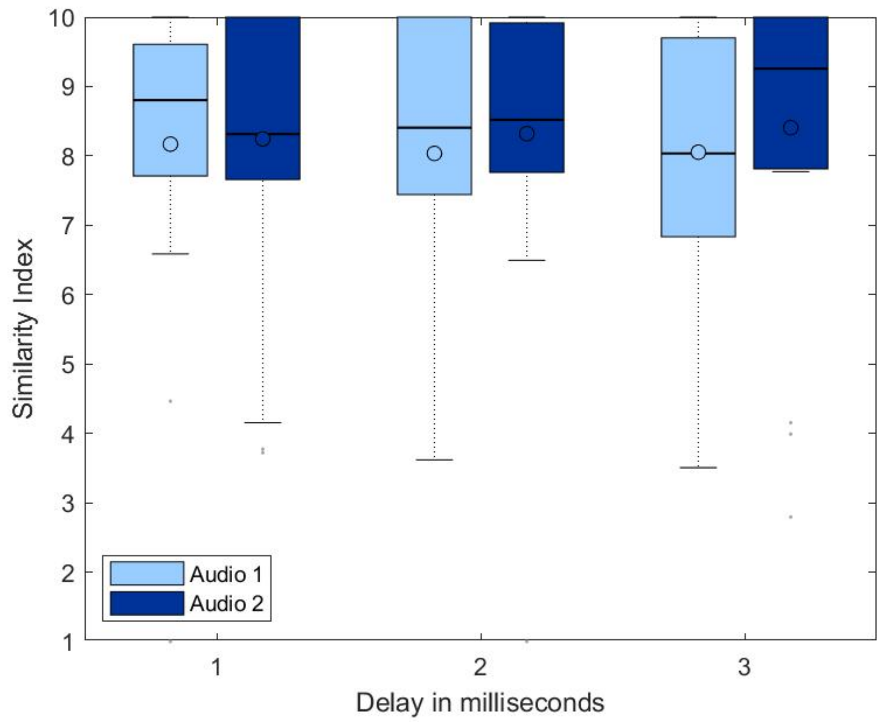


Figure 5.13: Test II Similarity Result for 60 Hz

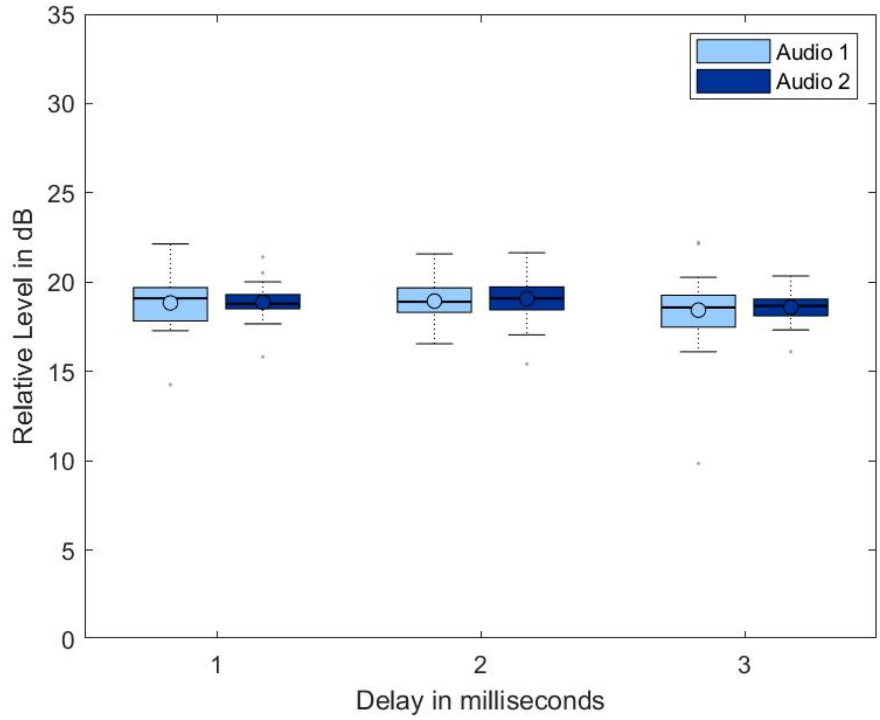


Figure 5.14: Test II Gain Result for 90 Hz

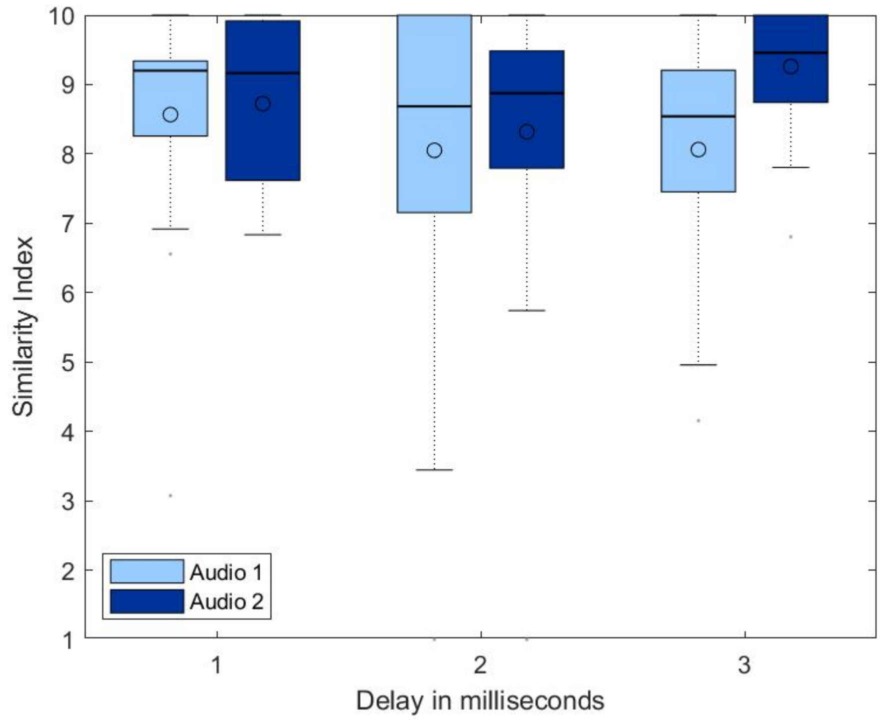


Figure 5.15: Test II Similarity Result for 90 Hz

5.2.3 Test III

The Results for Test III show that the gain level required to reduce the effect of the resonance is about 19dB. As expected, the variance in Audio 1 for 35 Hz is the highest among other frequencies due to the resonance level being close to its perceptual threshold.

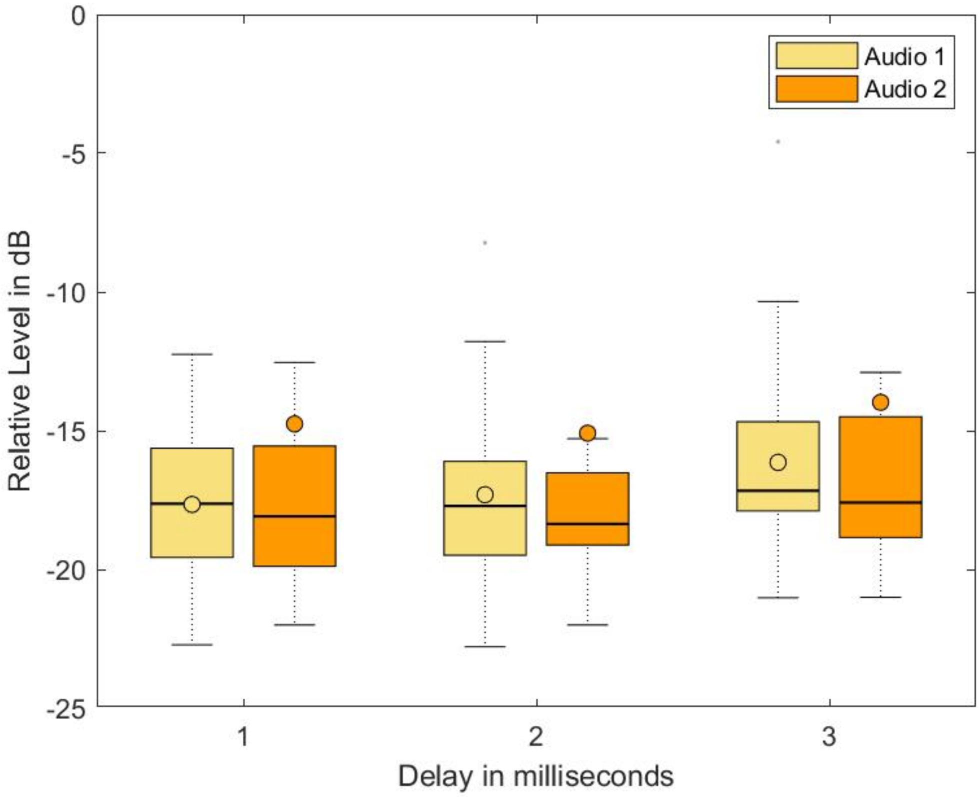


Figure 5.16: Test III Gain Result for 35 Hz

Unlike the results in Test II, it is expected to observe a high variance in similarity index for both audio samples. The reason being, the peak filter used to cut the resonance effect can easily be misjudged in level setting. There is an ambiguity of doubt on the exact level setting needed to reduce the resonance effect. There is a high chance of over setting the level needed, and this can affect the similarity. This is an obvious observation seen in all cases.

A key factor to be noted in Figure 5.17, 5.19 and 5.21 is the variabiliy in the similarity index. A clear indication in this case is that, although the subjects were able to match the reference audio in terms of gain level, the peak filter had an adverse effect in the quality and timbre of the audio, and much more prevalent in the case for Audio 1. The high variance in the results does suggest how much for each subject the audio is similar.

A possible reason for this high variance in similarity may be attributed to the fact low Quality factor values have adverse effects on all frequencies around the centre frequency of a peak filter. Determining the correct loudness may be tricky, especially if the audio contains a lot of transients. Since the clarity in audio 2 is high enough to perceive the correct loudness (as seen in figures 5.16, 5.18 and 5.20), the subjects were able to judge the timbre and quality of the audio, as compared to audio 1. If the subjects have perceived the wrong loudness, even as mucha as 3 dB, this would be enough to affect the surrounding frequencies, and thereby making the perceived audio less similar to the original.

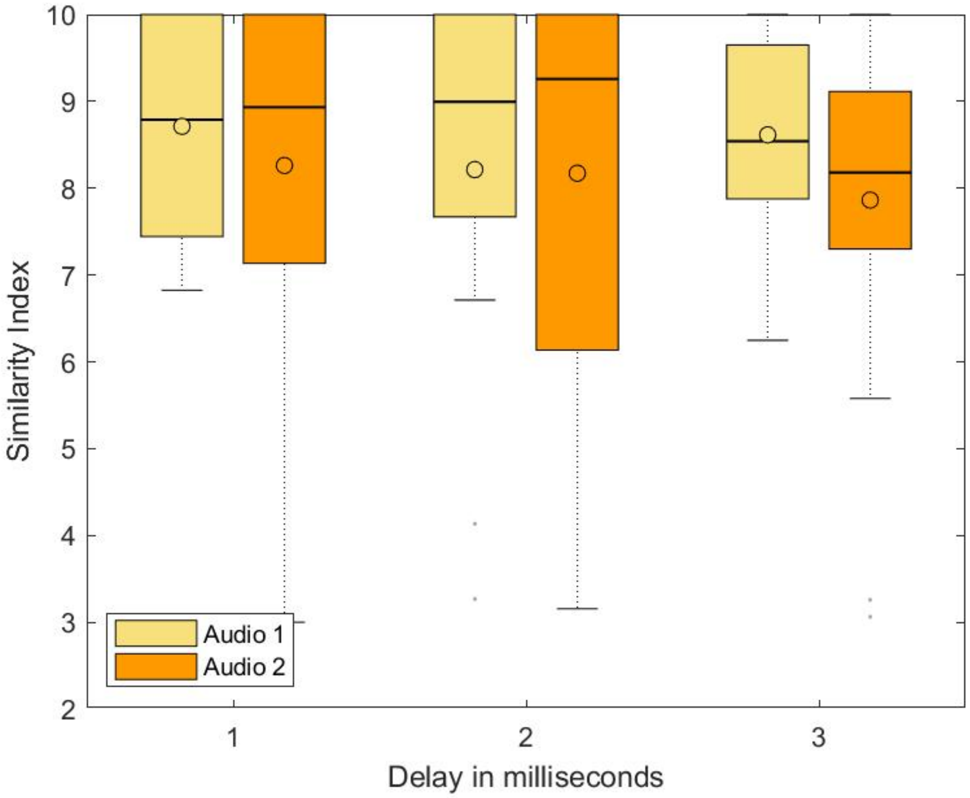


Figure 5.17: Test III Similarity Result for 35 Hz

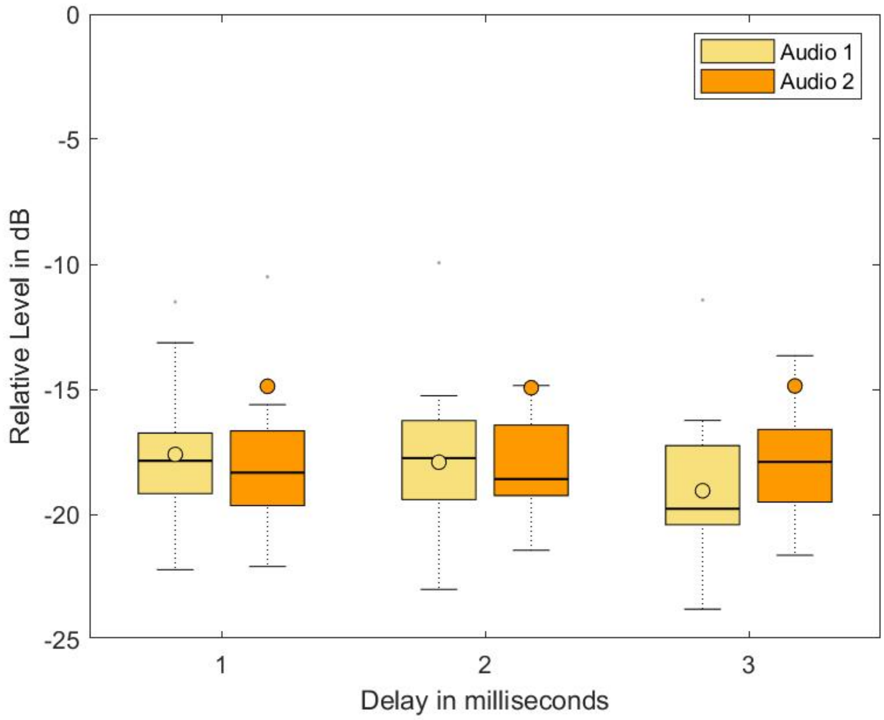


Figure 5.18: Test III Gain Result for 60 Hz

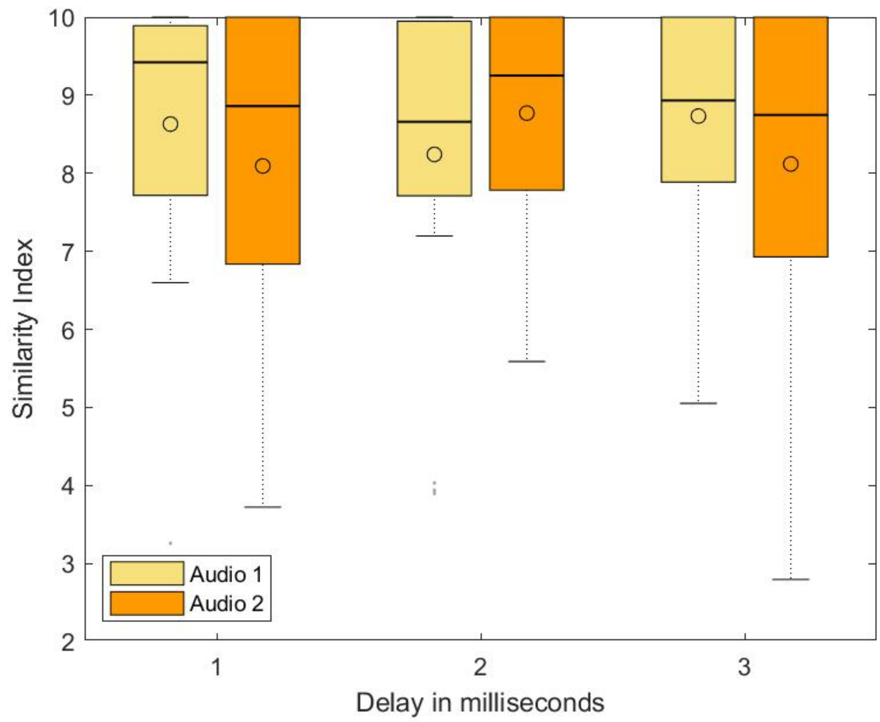


Figure 5.19: Test III Similarity Result for 60 Hz

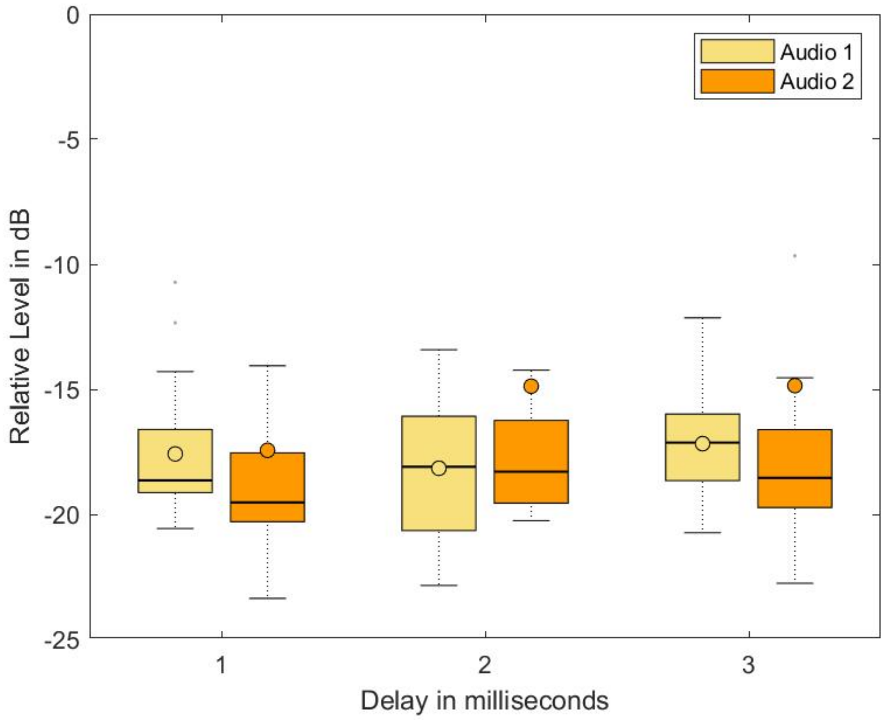


Figure 5.20: Test III Gain Result for 90 Hz

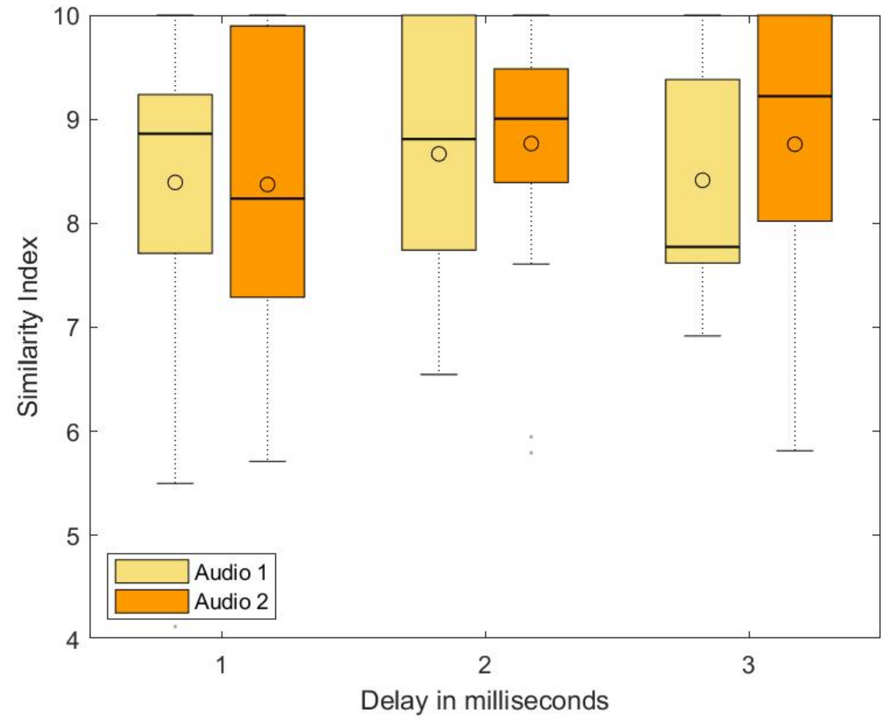


Figure 5.21: Test III Similarity Result for 90 Hz

5.3 Overview and Discussion on the Listening Test

The listening test results show a good agreement to what is expected. In Test I, the estimated perceptual threshold show the minimum level required to be able to perceive the low frequency resonances. As indicated in [2], this threshold level corresponds to a resonator's steady state level, L_r. The audibility of a resonance depends on the level difference between a resonance's maximum level and the system magnitude level, termed as ΔL. According to the author in [2], if ΔL is lower than L_r, the resonance will not be audible. In the case of a high end loudspeaker like the Genelec 8020B, whose level difference happens to be approximately 10 dB with a resonance frequency at 43 Hz, the audibility of the resonance will depend on the type of audio. If for example, audio 2 was used, the 43 Hz may correspond to a level of 5 dB, indicating that the resonance will be audible. In the case of a sub woofer like the Neumann KH 805, the level difference is much higher. This gives the necessary indication that in car loudspeakers, especially for sub woofers, for any additive resonances to be audible, their level differences have to be higher than the estimated steady state level, which varies with the audio used.

In the assumption that these additive resonances are audible, the question now arises as to how can one be able to measure the loudness of these additive resonances, and how should one be able to control the loudness. Conventional sound pressure level measurements are impractical because the resonance levels can be easily masked with any audio signal used.

This is where the purpose of Test II and Test III comes in. Although not so evident, the main purpose of fabricating these two tests is to create a model to emulate any resonance. With the rise of new car models and designs, a necessary method of verification would be needed to test the effect of the design of the mounting structures given to the car loudspeakers. Having to measure the resonance levels robustly is definitely impractical, time consuming and expensive. The methodology used to detect and estimate the resonances proved to show that it is possible to quantify and emulate any resonance. Combining the methodology and the listening test model can greatly enhance the ability for one to immediately estimate the effect of ringing, based on the design of the loudspeaker and mounting. In Test II, determining the level equivalence of the resonances can give one a range of severity levels to know when the ringing effect becomes severe, with respect to every type of audio used. Test III determines whether using a filter could minimize the resonance effect, while maintaining the spectral balance of the audio.

This indicates that one can create a loudness metric to the effect of ringing. Unfortunately, as mentioned in Chapter 3, this requires several tests that involve different environments and setups. However, this can pave way to the next steps of developing the metric.

5.4 Limitations

The methodologies used, as well as the listening test, come with many limitations. In the CSD method, unfortunately the resolution of the 3D plot comes at the cost of the resolution of the appodizing window. Inspite of being able to reduce the spectral leakage caused by the unit step window, the appodizing window causes a slight loss in power spectral density.
The CWT is very sensitive to change in scaling factor. Normally, to compute the CWT of a signal, a scaling factor is chosen to have a value in powers of 2, and the powers are proportional to the number of octaves needed to have adequate resolution, both in time and frequency. The CWT calculation was implemented manually, through MATLAB scripts, in spite of the program having the CWT functionality. This is because the resolution is fixated, without any freedom to vary paramters, such as the centre frequency, the bandwidth and the scaling factor. This would hinder the visual accuracy of the frequency data points, as the scaling factor would map the scaled wavelet for each frequency. Care is need here, in order to avoid massive divergence. As a consequence, the scaling factor value was not chosen to be in octaves, and thus hindered the resolution of the plot. With proper care and better scripting on calculating the CWT, the resolution can be improved further.
The Steiglitz-McBride method to recreate the impulse response can cause instability when creating the coefficients, depending on the length of the impulse response and the number of samples. Having a high number of samples can overload the CPU, to the point of crashing the system, if the RAM is overloaded. The resolution in this case, will become a challenge, and an adequate balance would be needed to have the best visualization and estimation of the resonances. Fortunately, since the focus is on the low-frequency region, downsampling the impulse response helps in reducing the coefficients needed to estimate the resonances. This would become problematic when estimating mid and high frequency resonances. There are alternatives to process long number coefficients, like parallel computing, which may be used to improve the computational efficiency, and also make more stable coefficients.
The statistical box plots used in the listening tests come at the limitation of the number of subjects that took part in the test. Some of the subjects have shown a high degree of deviation, whom had to be excluded from the calculation, as it had an impact on the result.
Originally, the prime focus on the usage of the methodology was on the measurement and validation of cavity resonances on a car's door panel. Due to the non availability of resources and time constraints, the focus for validation of the methodology was changed. The methodology would be more credible, if the chosen loudspeakers had been from an existing loudspeaker from a car. The Genelec and Neumann Loudspeakers have high fidelity and well dampened casings, that made the measurement and estimation of resonances a challenging task. In spite of this, since the methodology was well able to perform as it was supposed to be intended for, this could be considered a plus factor, considering that cavity resonances may be a bigger challenge to estimate.

6
Discussion, Limitations and Conclusion

6.1 Discussion

The listening test results show a good agreement to what is expected. In Test I, the estimated perceptual threshold show the minimum level required to be able to perceive the low frequency resonances. As indicated in [2], this threshold level corresponds to a resonator's steady state level, L_r. The audibility of a resonance depends on the level difference between a resonance's maximum level and the system magnitude level, termed as ΔL. According to the author in [2], if ΔL is lower than L_r, the resonance will not be audible. In the case of a high end loudspeaker like the Genelec 8020B, whose level difference happens to be approximately 10 dB with a resonance frequency at 43 Hz, the audibility of the resonance will depend on the type of audio. If for example, audio 2 was used, the 43 Hz may correspond to a level of 5 dB, indicating that the resonance will be audible. In the case of a sub woofer like the Neumann KH 805, the level difference is much higher. This gives the necessary indication that in car loudspeakers, especially for sub woofers, for any additive resonances to be audible, their level differences have to be higher than the estimated steady state level, which varies with the audio used.

In the assumption that these additive resonances are audible, the question now arises as to how can one be able to measure the loudness of these additive resonances, and how should one be able to control the loudness. Conventional sound pressure level measurements are impractical because the resonance levels can be easily masked with any audio signal used.

This is where the purpose of Test II and Test III comes in. Although not so evident, the main purpose of fabricating these two tests is to create a model to emulate any resonance. With the rise of new car models and designs, a necessary method of verification would be needed to test the effect of the design of the mounting structures given to the car loudspeakers. Having to measure the resonance levels robustly is definitely impractical, time consuming and expensive. The methodology used to detect and estimate the resonances proved to show that it is possible to quantify and emulate any resonance. Combining the methodology and the listening test model can greatly enhance the ability for one to immediately estimate the effect of ringing, based on the design of the loudspeaker and mounting. In Test II, determining the level equivalence of the resonances can give one a range of severity levels to know when the ringing effect becomes severe, with respect to every type of audio used. Test III determines whether using a filter could minimize the resonance effect, while maintaining the spectral balance of the audio.

This indicates that one can create a loudness metric to the effect of ringing. Unfortunately, as mentioned in Chapter 3, this requires several tests that involve different environments and setups. However, this can pave way to the next steps of developing the metric.

6.2 Limitations

The methodologies used, as well as the listening test, come with many limitations. In the CSD method, unfortunately the resolution of the 3D plot comes at the cost of the resolution of the appodizing window. Inspite of being able to reduce the spectral leakage caused by the unit step window, the appodizing window causes a slight loss in power spectral density.

The CWT is very sensitive to change in scaling factor. Normally, to compute the CWT of a signal, a scaling factor is chosen to have a value in powers of 2, and the powers are proportional to the number of octaves needed to have adequate resolution, both in time and frequency. The CWT calculation was implemented manually, through MATLAB scripts, in spite of the program having the CWT functionality. This is because the resolution is fixated, without any freedom to vary paramters, such as the centre frequency, the bandwidth and the scaling factor. This would hinder the visual accuracy of the frequency data points, as the scaling factor would map the scaled wavelet for each frequency. Care is need here, in order to avoid massive divergence. As a consequence, the scaling factor value was not chosen to be in octaves, and thus hindered the resolution of the plot. With proper care and better scripting on calculating the CWT, the resolution can be improved further.

The Steiglitz-McBride method to recreate the impulse response can cause instability when creating the coefficients, depending on the length of the impulse response and the number of samples. Having a high number of samples can overload the CPU, to the point of crashing the system, if the RAM is overloaded. The resolution in this case, will become a challenge, and an adequate balance would be needed to have the best visualization and estimation of the resonances. Fortunately, since the focus is on the low-frequency region, downsampling the impulse response helps in reducing the coefficients needed to estimate the resonances. This would become problematic when estimating mid and high frequency resonances. There are alternatives to process long number coefficients, like parallel computing, which may be used to improve the computational efficiency, and also make more stable coefficients.

The statistical box plots used in the listening tests come at the limitation of the number of subjects that took part in the test. Some of the subjects have shown a high degree of deviation, whom had to be excluded from the calculation, as it had an impact on the result.

Originally, the prime focus on the usage of the methodology was on the measurement and validation of cavity resonances on a car's door panel. Due to the non availability of resources and time constraints, the focus for validation of the methodology was changed. The methodology would be more credible, if the chosen loudspeakers had been from an existing loudspeaker from a car. The Genelec and Neumann Loudspeakers have high fidelity and well dampened casings, that made the measurement and estimation of resonances a challenging task. In spite of this, since the methodology was well able to perform as it was supposed to be intended for, this could be considered a plus factor, considering that cavity resonances may be a bigger challenge to estimate.

6.3 Conclusion

The results from both the estimation of resonances and the listening tests show a good agreement, giving the indication that the methodology can be used to detect cavity resonances. The combination of the CSD, CWT and System Identification methods can pin point the exact frequency of the resonance as well as determine its strength through the Quality factor. In addition, the perception of these resonances can be estimated using the models used in the listening test.

Due to the high variance in the perceptual test results, it is necessary to have a higher number of subjects in order to have an in-depth and conclusive result. This would greatly enhance the perceptual model, and make a better evaluation on the audibility of the resonances. The methodology for detection and estimation of resonances will serve well for detecting cavity and other mechanical resonances in a car for low to mid frequencies, if speed and efficiency is a requirement.

A further research on these methods may be pursued, by quantitatively estimating the loudness of the resonances, through a combination of existing algorithms that estimate the loudness. In addition, other psychoacoustical parameters such as roughness and sharpness can indicate the nature of resonances. This research can further be extended, by utilizing the power of machine learning, which can be used to predict the perception of resonances in terms of psychoacoustic parameters.

Bibliography

[1] Floyd E. Toole, Sean E. Olive (1988) The modification of Timbre by Resonances: Perception and Measurements. Journal of Audio Engineering Society, vol. 32, 122 - 142

[2] Ivo Mateljan, Heinrich Weber, Ante Doric, (2007) Detection of Audible Resonances, 3rd Congress of the Alps Adria Acoustics Association.

[3] Jacob Dyreby, Sylvain Choisel,(2007) Equalization of loudspeaker resonances using second-order filters based on spatially distributed impulse response measurements, 123rd Convention Audio Engineering Society, New York

[4] Shelley Uprichard, Sylvain Choisel,(2008) The Influence of Acoustic Environment on the Threshold of Audibility of Loudspeaker Resonances, 125th Convention Audio Engineering Society, San Fransisco

[5] Ethan Winer, (2012) Chapter 1 - Audio Basics, The Audio Expert, 3 - 39.

[6] Ethan Winer, (2012) Chapter 3 - Hearing, Perception and Artifact Audibility, The Audio Expert, 65 - 104.

[7] Nikos.(2019) A Living Room for the Evaluation of multiple auditory scenes, Master Thesis, Chalmers University of Technology

[8] John D.Bunton, Richard H. Small (1982) Cumulative Spectra, Tone Bursts and Apodization, Journal of Audio Engineering Society, vol.30, No.6

[9] S.J. Loudritis, (2005) Decomposition of Impulse Responses Using Complex Wavelets, Journal of Audio Engineering Society, vol. 53, No.9, 796 - 811

[10] D.B Keele, (1999) Time-Frequency Display of Electroacoustic Data using Cycle-Octave Wavelet Transforms, Audio of Engineering Society 99th Convention, New York

[11] K.Steiglitz, L.E. McBride,(1965) A Technique for Identification of Linear Systems, IEEE Trans. Automation Control, vol. AC-10, 461-464

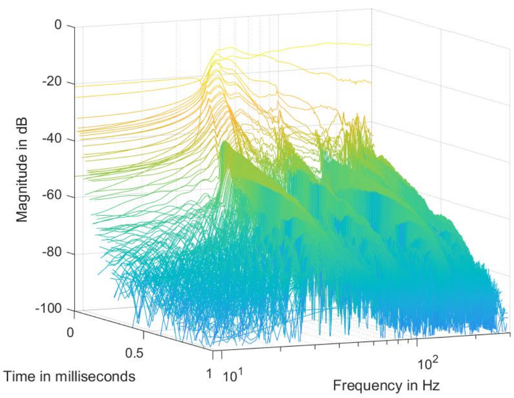
A

Appendix

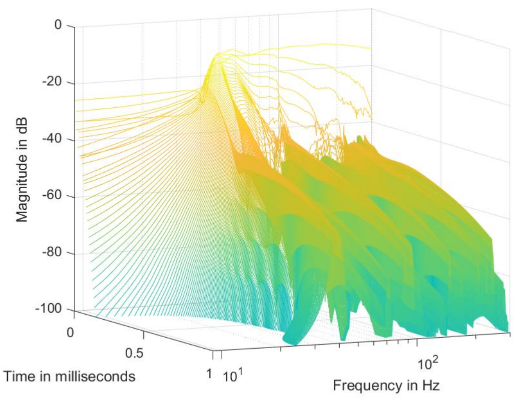
Estimation of Resonances

The following plots represent the estimated resonances from all the three methodologies used: Cummulative Sepctral Decay, Continuous Wavelet Transform and System Identification, for the loudspeakers Genelec 8020B and Genelec 7050B.

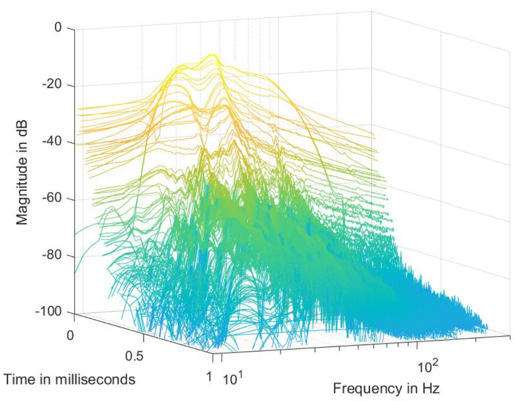
A.0.1 Cumulative Spectral Decay (CSD)



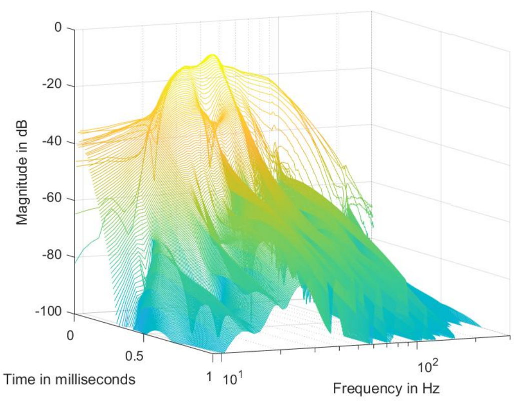
(a) Genelec 8020B-No Appodization



(b) Genelec 8020B-Appodization



(c) Genelec 7050B-No Appodization

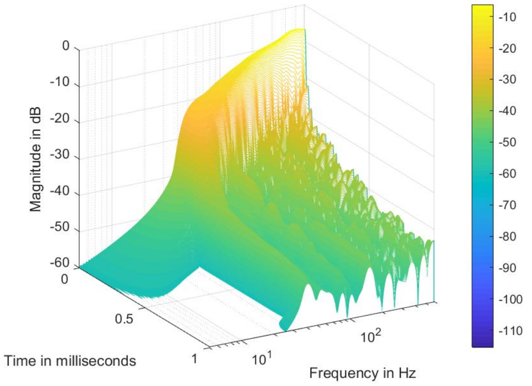


(d) Genelec 7050B-Appodization

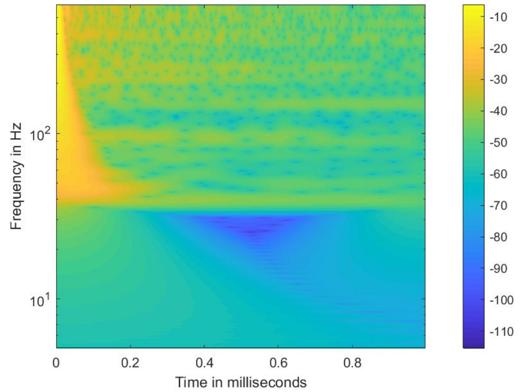
Figure A.1: Cummulative Spectral Decay Plot of Genelec 8020B and Genelec 7050B

A.0.2 Continuous Wavelet Transform (CWT)

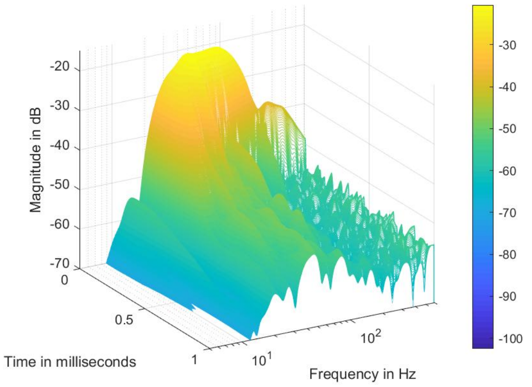
A.0.2.1 3D and 2D Magnitude Plot:



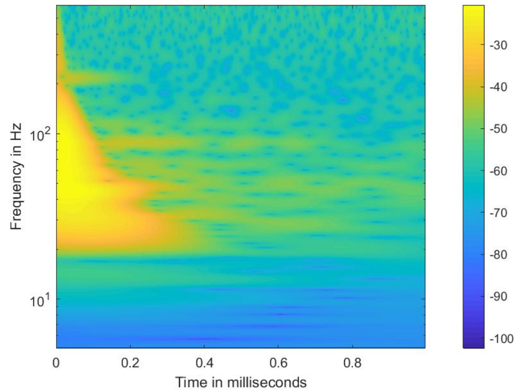
(a) Genelec 8020B - 3D View



(b) Genelec 8020B - 2D Top View



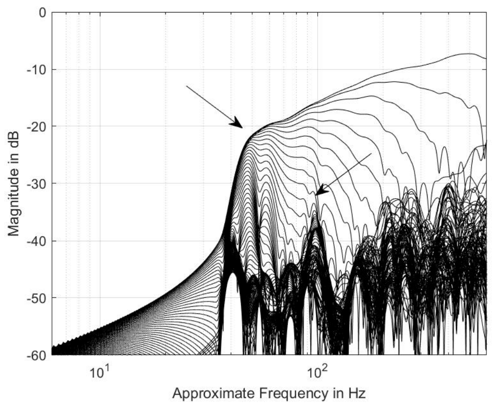
(c) Genelec 7050B - 3D View



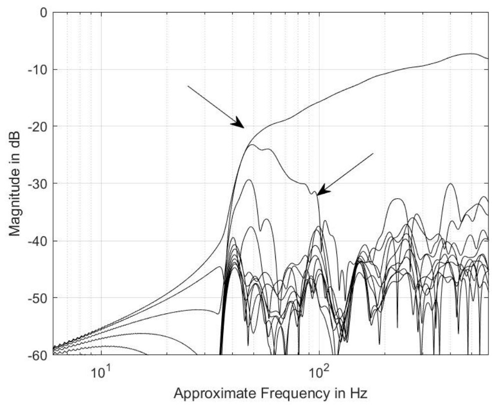
(d) Genelec 7050B - 2D Top View

Figure A.2: Continuous Wavelet Transform - Magnitude Plot of Genelec 8020B and Genelec 7050B

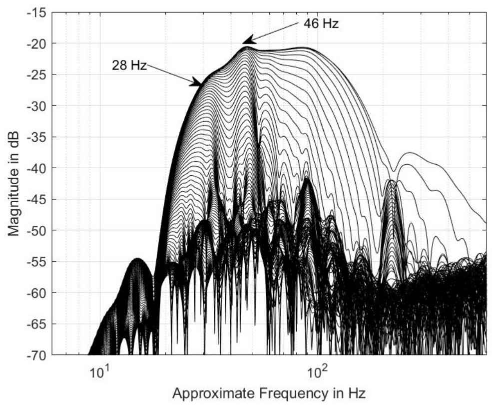
A.0.2.2 2D FrontView, Full and Half Slices:



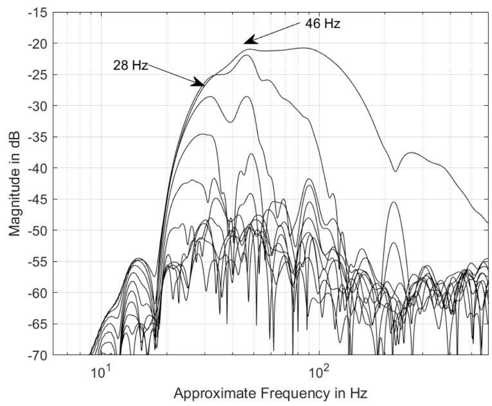
(a) Genelec 8020B - 2D Full Slice View



(b) Genelec 8020B - 2D Half Slice View



(c) Genelec 7050B - 2D Full Slice View

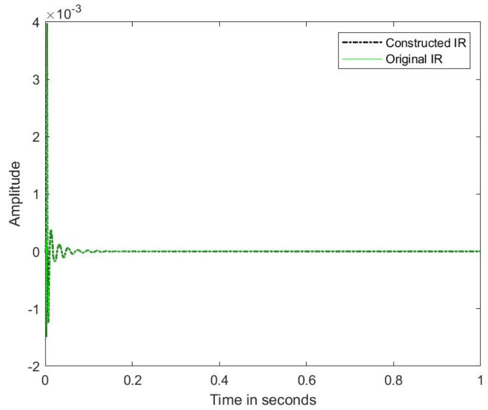


(d) Genelec 7050B - 2D Half Slice View

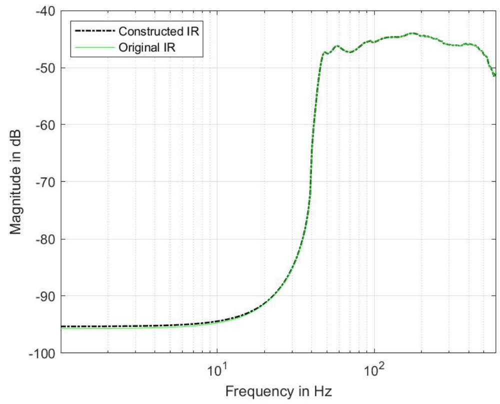
Figure A.3: Continuous Wavelet Transform - 2D Front View and location of Resonances of Genelec 8020B and Genelec 7050B

A.0.3 System Identification Method (Steiglitz - McBride)

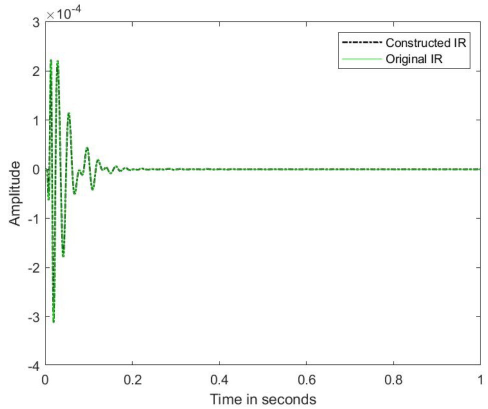
A.0.3.1 Full Impulse Response and Frequency Response Reconstruction:



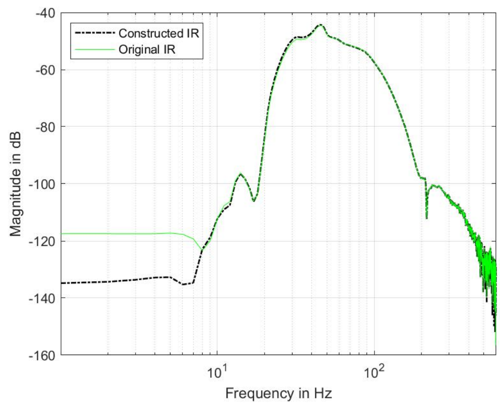
(a) Genelec 8020B - Original and Reconstructed Impulse Response



(b) Genelec 8020B - Original and Reconstructed Frequency Response



(c) Genelec 7050B - Original and Reconstructed Impulse Response

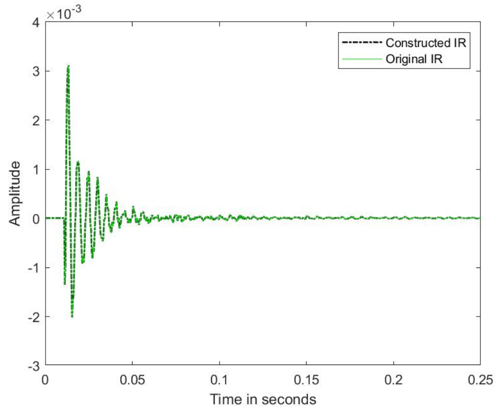


(d) Genelec 7050B - Original and Reconstructed Frequency Response

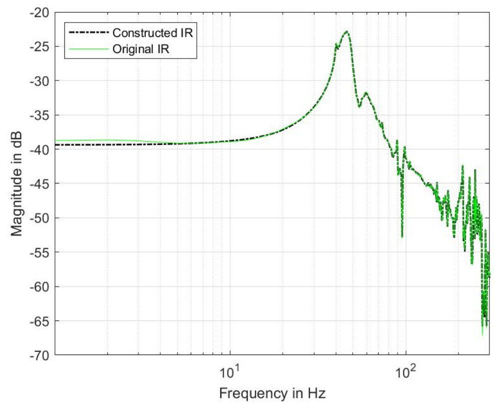
Figure A.4: Impulse Response, Frequency Response Reconstruction of Genelec 8020B and Genelec 7050B

A.0.3.2 Sliced Impulse Response and Frequency Response Reconstruction:

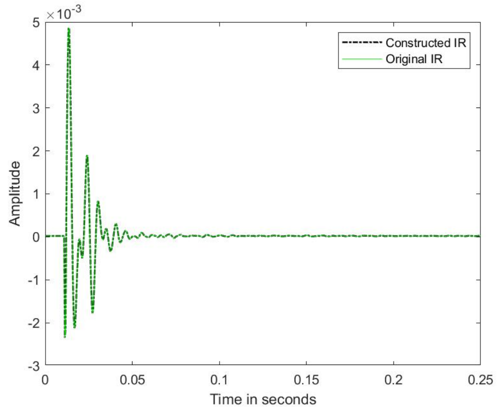
As explained in Chapter 4, the sliced IR is chosen from the CSD calculation at an arbitrary time value, in order to expose the resonance of the loudspeaker.



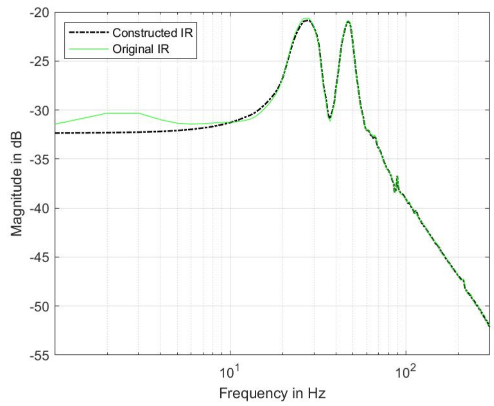
(a) Genelec 8020B - Original and Reconstructed sliced Impulse Response



(b) Genelec 8020B - Original and Reconstructed sliced Frequency Response



(c) Genelec 7050B - Original and Reconstructed sliced Impulse Response



(d) Genelec 7050B - Original and Reconstructed sliced Frequency Response

Figure A.5: Impulse Response, Frequency Response Reconstruction of Genelec 8020B and Genelec 7050B