POLITECNICO DI MILANO
Corso di Laurea Magistrale in Ingegneria Informatica
Dipartimento di Elettronica e Informazione

DSP Analysis and Psychoacoustic Testing on the Perception of Low Frequency Transients and Quality Degradation in Small Rooms

Supervisor: Prof. Giuseppe Bertuccio

External Supervisors: Ing. Lorenzo Rizzi
Ing. Gabriele Ghelfi

Master Graduation Thesis of:
Federico Ascari
Student ID: 815686

Academic Year 2014-2015

Contents

List of Figures

List of Tables

DSP Analysis and Psychoacoustic Testing on the Perception of Low Frequency Transients and Quality Degradation in Small Rooms

Abstract

The low-frequency behavior of small rooms has been studied thoroughly from a theoretical point of view in the past, but the psychoacoustic confirmation of such analyses is not complete. This research aims at bridging the gap between the theoretical analysis and the perceptual world, also introducing new theories that would explain the psychoacoustic behavior.

Nowadays, with the rise of new affordable technologies, it's becoming more and more frequent to work on music in small rooms, the so-called "home studios". It is known that small rooms, however, suffer from low-frequency problems caused by resonances and room modes. As a matter of fact, such ambients are non-sabinian and the modal region extends up high in the human hearing range, shaping and colouring the sound emitted by the speakers before arriving to the listener's ears.

The phenomenon of room modes is well known, but what is not well known yet is the specific perceptual effects that room modes, and, more in general, all problems in the lower range of the frequency response, have on listeners. An initial work in this field with these exact conditions was Michele Ferroni's thesis work[1], developed in collaboration with acoustical engineering studio Suono e Vita[2], who made a first step towards analyzing these phenomena.

In this research, Ferroni's results are the starting point of more advanced analyses, carried out using an expanded version of the Acoustic Quality Test originally developed by I. Adami and F. Liberatore[3], and later developed by A. Farina[4]. This advanced tool allows to analyze the temporal evolution and behavior of each specific frequency after the convolution with a room's impulse response.

This thesis work, developed in collaboration with Suono e Vita, studies the psychoacoustic perception of the effects introduced by resonance modes, with particular focus on the perception of levels for short notes and the degradation of precision and definition of the sound. First of all, a frequency domain and temporal domain analysis of eight impulse responses, acquired in small listening rooms with volume between 30 and 55 m^3, was carried out. The problematic behavior of such rooms, related to the presence of frequency response peaks and valleys, and frequencies with long decaying tails, were analyzed through two psychoacoustic tests in order to understand if these problems would be perceived by listeners.

In particular, the research focused on the importance of transient energetic phenomena regarding the perception of the level of short sounds (< 600 ms) at low frequency (between 20 and 300 Hz). Temporal parameters of the room such as Decay Time were analyzed, and two new parameters, called "Room Slowness" and "Room Inertia", have been defined. These parameters proved to be effective in quantifying the degradation of precision and definition of the sound after the convolution with an impulse response.

From the results, it appears that, for short sounds, the classic frequency response is not so significative regarding the perception of volume. Instead, the curve called "Overshoot Response" is proposed as a psychoacoustically more accurate counter-

part for only short sounds. Furthermore, the precision and definition of sounds is heavily modified by the convolution with the room impulse response and it appears to be strictly correlated with the temporal parameters of the room.

The results of this research offer important insights in the field of psychoacoustics and will allow to develop new criteria for the acoustic design of small rooms used for critical listening.

Notes

[1]M. Ferroni - "Evaluation and Psychoacoustic Validation of Techniques for the Analysis of Low Frequency Resonance Modes in Real Small Rooms", Master graduation thesis, Politecnico di Milano (2015)

[2]Suono e Vita, via Cavour 18, Lecco (LC), Italy

[3]I. Adami, F. Liberatore - "La messa a punto del sistema Diffusori - Ambiente", Acustica Applicata srl, Lucca, Italy

[4]A. Farina et al. - "AQT - A New Objective Measurement Of The Acoustical Quality of Sound Reproduction in Small Compartments", Audio Engineering Society, 110th convention, Amsterdam (2001)

Analisi DSP e Test Psicoacustici sulla Percezione di Transienti a Bassa Frequenza e sulla Degradazione di Qualità Percepita nelle Stanze di Dimensioni Ridotte

Sommario

Il comportamento acustico di stanze di dimensioni ridotte (caratterizzate da dimensioni comparabili alla lunghezza d'onda dell'onda sonora considerata) sollecitate da segnali sonori a bassa frequenza è stato studiato dal punto di vista teorico in passato, ma la conferma sul piano psicoacustico di tali analisi non è completa. Questa ricerca punta a ridurre la distanza tra le analisi teoriche e il mondo percettivo, introducendo nuove teorie che potrebbero spiegare alcuni fenomeni percettivi.

Al giorno d'oggi, con l'aumento di nuove tecnologie dai prezzi ridotti, sempre più frequentemente si lavora a produzioni musicali in stanze di dimensioni ridotte, i cosiddetti "home studios". E' risaputo, però, che ambienti piccoli soffrono di problemi di risposta alle basse frequenze causati dai modi di risonanza. Infatti, tali ambienti sono considerati non sabiniani e la regione modale si estende nella regione uditiva umana, colorando il suono emesso dagli speaker e modificando il suono percepito dall'orecchio dell'ascoltatore rispetto a quello emesso dalla sorgente.

Il fenomeno delle risonanze è ben conosciuto, ma ciò che non è ancora del tutto certo è in che modo, nello specifico, questo fenomeno, e più in generale, i problemi a bassa frequenza, si ripercuotono nella percezione degli ascoltatori su suoni che contengono transienti, come le note musicali. Un lavoro iniziale con le stesse premesse della presente ricerca è stato svolto da Michele Ferroni [1], per la sua tesi di Laurea Magistrale sviluppata in collaborazione con lo studio di ingegneria acustica Suono e Vita [2].

In questa ricerca, i risultati di Ferroni sono il punto di partenza di analisi più avanzate svolte utilizzando una versione estesa dell' Acoustic Quality Test, in origine ideato da I. Adami e F. Liberatore [3], poi sviluppato ulteriormente da A. Farina [4]. Questo strumento permette di analizzare l'evoluzione temporale e il comportamento di specifiche frequenze dopo la convoluzione con una risposta all'impulso di una stanza.

Questo lavoro di tesi, svolto presso Suono e Vita, si colloca nel suddetto ambito di ricerca ed ha riguardato la percezione psicoacustica degli effetti introdotti dai modi di risonanza, con particolare riguardo alla modifica della percezione dei volumi per suoni brevi e alla perdita di definizione del suono. In primo luogo è stata svolta una analisi nel dominio frequenziale e temporale di otto risposte all'impulso acquisite sperimentalmente in alcune stanze utilizzate per ascolto musicale, con volume compreso tra 30 e 55 m^3. I problemi emersi, legati alla presenza di alti picchi e profonde valli nella risposta in frequenza, e specifiche frequenze con lunghi tempi di decadimento, sono stati analizzati attraverso due test psicoacustici con il fine di capire se tali problemi sarebbero stati percepiti dagli ascoltatori.

In particolare, la ricerca è stata mirata all'analisi dell'importanza di fenomeni energetici di tipo transiente nella percezione dei livelli di ascolto di note brevi (< 600 ms) a bassa frequenza (tra 20 e 300 Hz). E' stato condotto uno studio su parametri temporali della stanza quali il tempo di decadimento e vengono proposti due nuovi

parametri, chiamati "Room Slowness" e "Room Inertia", che si sono rivelati efficaci per quantificare la perdita di precisione e definizione del suono dopo la convoluzione con una risposta all'impulso.

Dai risultati emerge che, per suoni brevi, la classica risposta in frequenza appare non essere particolarmente significativa riguardo alla percezione del volume sonoro. Invece, la curva chiamata "Overshoot Response" viene proposta come alternativa più appropriata dal punto di vista psicoacustico per suoni brevi a bassa frequenza. Inoltre, la percezione di precisione e definizione di ciascun suono è pesantemente modificata dall'interazione della stanza in seguito alla convoluzione con la risposta all'impulso, e sembra essere strettamente correlata ai parametri temporali della stanza stessa.

I risultati di questo studio aprono nuove prospettive nell'ambito della ricerca psicoacustica e permetteranno di fissare nuovi criteri nella progettazione acustica di stanze di dimensioni ridotte per ascolto critico.

Notes

[1]M. Ferroni - "Evaluation and Psychoacoustic Validation of Techniques for the Analysis of Low Frequency Resonance Modes in Real Small Rooms", Master graduation thesis, Politecnico di Milano (2015)

[2]Suono e Vita, via Cavour 18, Lecco (LC), Italia

[3]I. Adami, F. Liberatore - "La messa a punto del sistema Diffusori - Ambiente", Acustica Applicata srl, Lucca, Italia

[4]A. Farina et al. - "AQT - A New Objective Measurement Of The Acoustical Quality of Sound Reproduction in Small Compartments", Audio Engineering Society, 110th convention, Amsterdam (2001)

Dedication

To my family and Carol, for the support during these years.

Acknowledgements

I would like to thank Ing. Lorenzo Rizzi and Ing. Gabriele Ghelfi at Suono e Vita and Prof. Giuseppe Bertuccio at Politecnico di Milano.

Chapter 1

Introduction and Motivation

1.1 Introduction

In the last few years, the music recording industry faced a major revolution. Thanks to the countless new technologies that allow anyone to record and produce their music, and because of the availabilty and low cost of such software, the majority of music production shifted from big, iconic, acoustically treated recording studios, to home studios. In fact, even some hits have been produced in home studios or small studios. These rooms are, most of the times, equipped with the bare minimum tools necessary to record music. Sometimes, these rooms are not even dedicated to this use, to the point where no acoustic treatment is present.

While the comfort and convenience of this solution is undeniable, the conditions of these rooms put serious limits to the potential final quality of the product. In fact, small rooms have serious low frequency problems that can shift the perception of the sound, leading the mixing engineer to take unappropriate decisions regarding his work. This, when music is released, brings to poor translation of the song to other playback systems.

In particular, the most problematic phenomenon that appears in slow rooms is the one of resonances and room modes, that is, the increase or decrease in level of specific areas in the frequency response caused by the buildup or cancellation of acoustic waves whose wavelength is proportional to each of the room's dimensions. This is especially problematic for low frequencies since, above the so called "Schroeder Frequency", the sound field in a room can be considered statistically. While this phenomenon happens in all closed environments, it is particularly problematic in small rooms as the Schroeder Frequency is higher in the frequency response, therefore leaving a modal region which extends well into the human listening range, which goes from 20 to 20.000 Hz and corresponds to about 10 octaves. A more in-depth analysis of the acoustic of small rooms will be featured in the next chapters. It is clear that this problem needs to be addressed, because it is very difficult to treat small rooms in order to be efficient at low frequencies, both for budget reasons and because such treatment would require lots of space.

It is very important to understand how much these phenomena are perceived by the listener, even though not much regarding this topic is present in the literature.

A first work aimed at addressing the specific problem of the perception of resonances in real, small rooms and their perceived effects was done by Michele Ferroni in his thesis work at Polytechnic University Of Milan ([37]), developed in Conjunc-

tion with Suono e Vita, an italian acoustical engineering studio. Previously, similar works have been developed, but focusing on slightly different topics. Fazenda, Avis and Davies studied the perception limen of Q factor at low frequencies with virtual room models ([14]), and the dependance of the perception of room modes on room aspect ratios ([13]). Halmrast ([25]) studied the perception of the timbre and strenght in a small real room before and after treatment, while Salava ([42]) wrote an article regarding the audibility of imperfections at low frequencies. While this article does not provide definitive results, it offers interesting considerations that will be kept in mind during the development of this work. Hill and Hawksford ([26]) studied the low-frequency temporal accuracy of small room sound reproduction, focusing on the interaction between frequency response and phase response and how different correction methods may benefit one more than the other, leaving certain behaviors uncorrected. Karjalainen et al. ([51] studied the perception of Temporal Decay of low-frequency room modes. More recently, Fazenda et al. ([9]) defined, through psychoacoustic tests performed with synthetic room models and both musical and non-musical test sounds, perception thresholds for the detection of effects of room modes as a function of decays.

The results and considerations contained in these works have been recalled in the next chapters.

This research starts on the results of [37] and uses the same real room impulse responses measured by Suono e Vita, developing further analysis with an advanced version of the Acoustic Quality Test originally introduced by Farina ([8]). Two psychoacoustic tests with 30 musicians and expert listeners each will be performed (on headphones) to address the validity of the analyses' results, focusing mainly on the importance of transient energetic behavior regarding the perception of levels for low frequency short sounds, and on the temporal parameters of the room regarding the perceived degradation of precision and definition.

1.2 Overview of the acoustics of Small Rooms and Resonance Modes

The definition of "small room" is somewhat ambigous. As a matter of fact, the dimensions of an environment have to be compared with the wavelength of each frequency in order to define the environment's size in acoustic terms: a "large size" is about 10 times the wavelength of the considered frequency ([22]). This introduces an important differentiation: since high frequencies have very little wavelengths, while low frequency can have waves which are long many meters (a 100 Hz frequency wavelength is 3.3 meters), most environments are "big" with respect to high frequencies and "small" with respect to low frequencies. In this research, all rooms had volume below 60 cube meters and almost all room dimensions were smaller than five meters.

The Schroeder frequency is the theoretical limit frequency above which the behavior of the room can be considered statistically, and the frequency response varies to a smaller degree with respect to the placement of speakers and microphone. The Schroeder frequency depends on the room's dimensions and on the decay time. Below this frequency instead, the so-called "modal region" begins, in which the frequency response is dominated by resonances and room modes that create buildups

and valleys in the frequency response, strongly modifying the listening experience in that room. Stationary waves, also called "standing waves", are generated by interference phenomena at frequencies where the wavelength follows a specific relation with the room's dimensions. They do not propagate energy in the room and the locations at which the amplitude is minimum are called nodes, while those where the amplitude is maximum are called antinodes. The presence of buildups, which cause peaks in the frequency domain, is generally strongly correlated to long decaying tails at those specific frequency and high Q values. Valleys instead are present when at the receiver point the energy is very low, and can be caused by destructive interference by the direct and reflected field ([8]) or lack of stationary waves.

The frequency response curve for rooms is actually a Loudspeaker - Room transfer function, since it contains both the frequency response of the speaker, and the room behavior for that specific combination of speaker placement and microphone placement. All of these variables are very important and the correct placement of speakers and of the listening point may partially correct the frequency response curve even before treating the room: avoiding to place the speakers and listening position at anti-nodal points of the frequencies whose wavelenght relates to the room's dimensions is a good start. The reader is invited to read more on the correct placement of speakers and listening position in small studios, and the design of recording studio, if interested ([39], [50]).

Classic room correction methods work poorly in this situation, because all of them, in order to be efficient, have to take up lots of space. This is almost always impossible in small rooms.

Chapter 2 will be focused towards recalling and explaining in details these and more room acoustic concepts.

1.3 Analysis of previous results

The following list contains the results of Ferroni's preliminary research ([37]).

- Higher amplitude peak doesn't necessarily correspond to higher perceived loudness. There isn't a direct link between peak amplitude and loudness perception.

- Constructive resonances at low frequencies can cause energization of a sound making it louder even if the amplitude peak is not particularly reinforced.

- Destructive effects due to resonances at low frequencies can decrease the energy of a sound making it less louder, decreasing body of the sound.

- Resonance modes can affect significantly the timbre of kick drum hits. In some cases amplitude peak of the kick can be shifted from the attack to the body of the sound, even though the resonance has to be very powerful to make this effect perceivable.

- Resonances seem to be more perceivable for sounds played by pitched instruments, with notes of sufficient duration.

- Duration of sounds plays a key role in the perception of resonance modes. It seems that a sound affected by resonances can be perceived more or less loud depending on its duration.

- If resonances are strong enough they may affect the perception of onsets for sounds repeated with sufficient speed, causing attacks to be less defined, damped or distorted.

- Resonance modes always introduce a degradation of listening quality, producing a not faithful sound playback and unbalancing frequency response of the environment.

- Human auditory system usually has problems to identify loudest or less loudest sound, when average amplitude differences are below 1/1.5 dB and sounds are frequentially near.

- The presence of instruments also at higher frequencies does not seem to mask problems due to resonance modes at low frequencies.

In particular, the fact that the frequency response curve is not always relatable to the perceived loudness, alongside the Overshoot concept introduced in [8], lead to the idea that the room's energetic transient behavior can play a major role in level perception. Also, the result about the different perception of resonances for pitched and unpitched instruments raises some questions about the nature of the spectrum of test sounds and the masking that they potentially cause in relation to low frequency problem. These, and all other results, were inspected through further questions in the new tests.

The following list contains some notable results by other researchers that were kept in mind during the development of this work.

- An initial work on the vocabulary of descriptors for low frequency perceived behavior was developed in [36]. In the same research, it is reported that modal behavior can be noticed above the Schroeder frequency, which depends both on the room dimensions and on the decay time. Reducing the decay time results in a clear perceptual improvement and seems to be more important than speaker placement. In the present research, the vocabulary used for the psychoacoustic test questions was the one developed in [37] through focused questions, since they were developed with the same rooms' impulse responses and type of audience (italian people) who would be featured in this research.

- With a similar scenario to the present research, in [35], 250 milliseconds tone bursts featured Response Envelopes that did not reach their actual steady state, whereas 600 milliseconds tone bursts' response envelopes did. For this reason, similar values will be used in order to carry out the analyses of this research.

- in [14], a detection limen for the Q factor regarding the perception of room modes has been proposed, indicating as 16 the value of Q below which room modes would go unnoticed.

- In the research regarding the dependance of perception of modal distribution and room aspect ratios, attempts to rank critical listening spaces based on modal distribution metrics were suggested to be highly misleading ([13]), since what is rated is not just the room but the interaction between the input stimulus and the room. In the present research, room rankings have been developed with the temporal parameters of the room rather than modal parameters, therefore providing a different scenario. This experiment has also been carried out with four different types of sound in order to try to minimize the effect of different source material.

- Ideal performance of a listening system can not be guaranteed only by reducing the magnitude error. The phase response should be addressed as well, trying to ideally minimize both magnitude and phase error, which are not always correlated ([26]). In [26], the authors hint at the fact that the system they are analyzing is minimum phase for frequencies below the Schroeder's frequency, possibly implying that the discrete modal region could be defined as the frequency band exhibiting minimum phase behavior (this was not confirmed as an official result, but just a hint for further developments).

- Below about 100 Hz, the human auditory perception of temporal details degrades and longer decays seem to go unnoticed if the magnitude response is well equalized ([51]).

- The results of listening tests using synthetic test signals are generally more consistent than those obtained with music test signals ([42]). Also, in [42], the author states that opinions still differ in many aspects of low frequency behavior, one of which regards wether gross irregularities of the Loudspeaker − Room Transfer Function are more audible than "rather slow build-ups and decays of low frequency room resonances". He states that his opinion is the first ones are more perceivable. In the present research, this has been examined and it seems that the slowness of decays is very relevant for short sounds, while for longer sounds the behavior of the LRTF is revealed and is more relevant.

- Musicians playing in a real room preferred it when the room was dampened, taming part of the resonances and reducing high frequency shimmering. In particular, it is suggested that trying to reduce shimmering in the high frequency range and room resonance in the bass could be more important when building a small room for music, if compared to a common reverberation time approach ([25]).

- Listening tests carried out with headphones and speakers provided consistent similarity and preference judgements. Furthermore, tests on headphones showed slightly better consistency with narrower confidence intervals regarding similarity and preference ratings ([52]).

- Psychoacoustic tests using both non-musical and musical signal, convolved with room models generated by auralization processes were performed in [9], deriving perception thresholds for the effects of resonance modes due to decay time. Results are in accordance with [51] and confirm that the choice of test samples is very important in psychoacoustic testing, because their content can

highly impact the perception of the phenomena under test. Thresholds computed with non musical stimuli appear to decrease with frequency until about 100 Hz, where they converge to around 0.2 s. Average thresholds measured with artificial stimuli are 0.9 s at 32Hz, 0.3s at 63Hz, 0.27s at 100Hz, 0.18s at 150Hz, and 0.17s at 200Hz. With music stimuli, results follow the same trend, even though an increase in the data variance hints at the fact that it is probably more difficult to detect decaying modal energy because of the temporal masking taking place in musical signals. Average thresholds measured with music stimuli are 0.51s at 63Hz, 0.3s at 125Hz, and 0.12s at 250Hz. With such signals, the effects of resonance modes are perceived both by the actual decaying tail and a change in timbre. It is important to note, however, that these values were generated with a synthesized room model impulse response, while this research focuses on the behavior of real rooms.

1.4 Developments and innovation of this research

The main innovation of this research with respect to the previous work is the use of the Acoustic Quality Test ([8]). In particular, an evolution of this algorithm was developed, introducing new features such as Decay Time computation, advanced temporal analysis, advanced Overshoot analysis, "Room Slowness" and "Room Inertia" computation.

This research is focused on the perception and effects of the presence of energetic transient behavior (called "Overshoots" at the beginning and end of the envelope of a frequency after the convolution with the room impulse response (called "Response Envelope" in the following), which happen on certain areas of the frequency response. This behavior was already hinted in [8], but the psychoacoustic confirmation of this phenomenon is missing from the body of literature.

Furthermore, the aim of this research is expanding the results obtained in [37] regarding the perceived loss of precision and definition that a sound undergoes when convolved with an impulse response. In order to do so, temporal metrics were introduced and analyzed, and a connection between their values and the perceived sound degradation was searched.

As already mentioned, Ferroni's work used the same impulse responses that were used in this research. In particular, while not being included in the final conclusions, at one point he mentioned that testers generally preferred room PRZ to room SGR when asked "which room allows for a more uniform listening experience". After analyzing the results, this question seemed ambigous because the concept of "uniform listening experience" is not well defined (the result was unexpected as well, because of the strange frequency response of room PRZ that will be shown in chapter 5). Therefore, this was the starting point to develop different questions in this research. In fact, questions were made in this research's second test regarding the perceived degradation of precision caused by each room, and the overall subjective preference of each room. This led to interesting results that will be analyzed in the next chapters.

1.5 Limitations of this research

As already stated, this research started from a phenomenological analysis of real room impulse responses. Therefore, in order to confirm the results obtained with this research, validation should be done by analyzing even more impulse responses and using also virtual room models, in order to change variables one at a time, and performing psychoacoustic tests on even more listeners.

Furthermore, the psychoacoustic tests were performed on headphones in order to test many rooms avoiding the influence of the one in which the test was performed. For this reason, test sounds were developed by convolving dry sound with the rooms' impulse responses (this is the reason why this research talks about the precision loss "after the convolution with an impulse response"). The behavior is expected to be the same ([52]) if the dry sound is played in the real room with source and receiver at the same positions that were used to measure the impulse response. However, psychoacoustic tests in real rooms using speakers are suggested in order to confirm these results.

Chapter 2

Fundamentals of Room Acoustics

In order to understand the following parts of this research, the main principles of room acoustics have to be recalled. These basics include sound propagation in closed space and the solution of the wave equation which results in the formula to compute room modes; the definition of frequency and phase response, which describe how a sound is transmitted from the source to the receiver; the Schroeder Frequency and the definition of Sabinian and non-Sabinian environments; the effects of reflections in closed spaces; finally, the definition of the most important acoustic parameters. There would be many more important principles to recall, therefore the reader is invited to delve deeper in the subject if he or she has an interested in understanding more profoundly the concepts of room acoustics. Some suggested reads are [33] , [34], [50], [39].

2.1 Overview of the propagation in closed spaces

There are many approaches that can be used to describe the propagation of sound in closed spaces. Among them, a common one is the formal solution of the wave equation. By applying boundary conditions that take into account the behavior of the sound on the surfaces of walls, the wave equation is solved, leading to a formula that is used to compute the eigenfrequencies of the closed space (also called "room modes"). This computation, however, is strictly related to the theoretical and ideal model of an empty and perfectly symmetric "shoe-box" space, providing results that can differ slightly from the real-life ones caused by the furniture, windows, acoustic treatment or wall material in real rooms. If rooms have different shapes, the room modes are harder to compute and no precise method exists.

A second approach is Geometrical acoustics, which thinks of the sound waves as if they were light rays, treating them with geometrical rules and studying their propagation based on the reflections along the walls that happen after the sound has been emitted by a source. This approach can be used if the wavelength is considered small with respect to the room, therefore it is of no use for this research, which deals only with low frequencies.

The third approach is the statistical one, which considers the energy balance between the direct and reflected field, viewing sound waves as small "energy packets" ([23]). This approach can be used only above the Schroeder frequency, therefore it won't be used in this work since the aim of this research is to study the temporal behavior of the modal region, which is, by definition, the one below the Schroeder

frequency.

In order to describe the sound propagation in closed spaces, it is necessary to start from the linearized wave equation, also known as Helmholtz equation, a second order differential equation used to describe the wave motion in terms of sound pressure with respect to space and time:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{2.1}$$

where p is the sound pressure measured in Pascal and c is the speed of sound (343 m/s). A complete and thorough derivation of this formula can be found in [33]. The fact that this formulation is linear implies the sound field generated by a sinusoidal source will feature a pressure that varies sinusoidally at all positions, and that linear superposition exists in this context ([23]), meaning that the total pressure at a certain point results from the superposition of the effects of all frequencies that constitute the sound wave.

2.2 Modes in Rectangular Rooms

In this research, the impulse response of real rooms with different shapes has been used for the analysis and psychoacoustic tests. While the rooms were not always simmetric or having a parallelepiped shape, the most general method to explain room modes is considering such rooms only. Of course, a certain degree of abstraction is present, and the room modes calculated with the final formula of this section will differ slightly in the real world because of the specific dimensions, furniture, acoustic treatment, irregularities of real rooms.

Considering a room with a parallelepiped shape, the calcution of normal modes is greatly simplified. The room's dimensions range from x = 0 to x = Lx in the x-axis, and, likewise, from y = 0 to y = Ly and from z = 0 to z = Lz respectively for the y and z-axis. Walls are assumed to be rigid, and the normal components of the particle velocity is equal to zero at the surface of the wall (boundary condition). The Helmholtz equation can be written, in cartesian coordinates, as:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} + k^2 p = 0 \tag{2.2}$$

where k is the wave number in angular notation.

The pressure can be separated in different components, each one depending on one dimension:

$$p(x, y, z) = p_x(x) \cdot p_y(y) \cdot p_z(z) \tag{2.3}$$

turning 2.2 into

$$\frac{\partial^2 p_x(x)}{\partial x^2}\cdot p_y(y)\cdot p_z(z) + \frac{\partial^2 p_y(y)}{\partial y^2}\cdot p_x(x)\cdot p_z(z) + \frac{\partial^2 p_z(z)}{\partial z^2}\cdot p_x(x)\cdot p_y(y) + k^2(p_x(x)\cdot p_y(y)\cdot p_z(z)) = 0 \tag{2.4}$$

If 2.4 is divided by $(p_x(x) \cdot p_y(y) \cdot p_z(z))$, the result is

$$\frac{1}{p_x(x)}\frac{\partial^2 p_x(x)}{\partial x^2} + \frac{1}{p_y(y)}\frac{\partial^2 p_y(y)}{\partial y^2} + \frac{1}{p_z(z)}\frac{\partial^2 p_z(z)}{\partial z^2} + k^2 = 0 \tag{2.5}$$

It is possible to see that each term in eq. 2.5 depends on one variable only, except for k^2 which is a constant. Since only one term depends on x, the term itself can not vary with x as the other members would have to change as well with x in order to keep the equation true. This is not possible, since the other members do not depend on x. Since this is a contradiction, every term in the equation must be constant, yelding:

$$\frac{1}{p_x(x)}\frac{\partial^2 p_x(x)}{\partial x^2} = -k_x^2 \tag{2.6}$$

$$\frac{1}{p_x(y)}\frac{\partial^2 p_y(y)}{\partial y^2} = -k_y^2 \tag{2.7}$$

$$\frac{1}{p_x(z)}\frac{\partial^2 p_z(z)}{\partial z^2} = -k_z^2 \tag{2.8}$$

where the constants k_x^2, k_y^2 and k_z^2 are related by

$$k_x^2 + k_y^2 + k_z^2 = k^2 = (\frac{2\pi f}{c})^2 = (\frac{\omega}{c})^2 \tag{2.9}$$

Equation 2.6 can be rewritten as:

$$\frac{\partial^2 p_x(x)}{\partial x^2} + k_x^2 * p_x(x) = 0 \tag{2.10}$$

For the sake of clarity, the following will regard only the development of eq. 2.10, but analogous considerations can be made also for 2.7 and 2.8. The solution of 2.10 is:

$$p_x(x) = A_1 \cdot cos(k_x x) + B_1 \cdot sin(k_x x) \tag{2.11}$$

where the constants A_1 and B_1 are used to adapt the solution to the boundary conditions, which are

$$\frac{\partial p_x(x)}{\partial p} = 0 \quad at \;\; x = 0 \quad and \quad x = Lx \tag{2.12}$$

since the particle velocity, which is the first deritvative of pressure, has to be set to zero at the boundaries.

By applying the boundary conditions, 2.12, to 2.11, it is possible to understand that B_1 has to be set to zero because its derivative, the cosine, is different from zero for arguments equal to 0 or Lx. Instead, the derivative of the cosine is equal to zero at $x = 0$. In order to satisfy 2.12, it also has to be equal to zero for $x = Lx$. To accomplish this, k_x Lx has to be a multiple of π. For this reason, the acceptable values for k_x are:

$$k_x = \frac{n_x \pi}{Lx} \tag{2.13}$$

where n_x is a non-negative integer. Combining this results with the analogous results obtained for 2.7 and 2.8, eingenvalues of the wave equation can be computed:

$$k_{n_x n_y n_z} = \pi\sqrt{(\frac{n_x}{Lx})^2 + (\frac{n_y}{Ly})^2 + (\frac{n_z}{Lz})^2} \tag{2.14}$$

The eigenfrequencies are:

$$f_n = \frac{\omega_n}{2\pi} = \frac{c}{2}\sqrt{(\frac{n_x}{Lx})^2 + (\frac{n_y}{Ly})^2 + (\frac{n_z}{Lz})^2} \tag{2.15}$$

where n_x, n_y, n_z are the indices of the room's dimensions. These frequencies are such that their velocity is minimum on the boundaries, and the pressure is maximum.

Finally, at each point in space, the pressure can be described by:

$$p_{n_x n_y n_z}(x, y, z) = C \cdot cos(\frac{n_x \pi x}{Lx}) \cdot cos(\frac{n_x \pi x}{Lx}) \cdot cos(\frac{n_x \pi x}{Lx}) \tag{2.16}$$

where C is a constant. This describes the pressure in a standing wave, and it is equal to zero each time one of the cosine functions equals zero. Of course, 2.16 is missing an exponential term that describes the temporal evolution of the pressure [33]. However, it is clear that the wave pattern does not advance in space, because the nodes are related to the room dimensions. Instead, pressure and velocity each vary in amplitude in some points in the room (with a maximum excursion on antinodes), and remain equal to zero in other points of the room (nodes).

Fig. 2.1 ([50]) represents the pressure of sound waves at the first order resonant frequency (subfigures a and b), at the second order resonant frequency (subfigure c), the velocity in the same case (subfigure d), which is zero at the boundary and maximum where the pressure is null, and a common way to represent the sound level (subfigure e).

By setting the indices values in 2.15 to non-negative integer numbers, different eigenfrequencies can be obtained. When only one of these numbers is different from zero, the mode is called Axial, since it is related only to one dimension of the room, and regards frequencies which resonate by bouncing on two opposite surfaces. These modes are usually the most energetic ones and most significative in shaping the room's amplitude response. When two of the indices are different from zero, the mode is called tangential and it regards a frequency which bounces on a plane, on four different surfaces. When all indices are different from zero the mode is called oblique and the resonant frequency bounces on all six walls of the room. This behavior is summarized in fig. 2.2.

The number of the indices describes the number of nodes in the wave (points where the pressure is equal to zero).

In particular,

- Axial modes

  The lowest resonant frequency of the room is the axial mode related to the bigger dimension.

  - x-axial modes, parallel to the x-axis ($n_x \neq 0$, $n_y = n_z = 0$)
  - y-axial modes, parallel to the y-axis ($n_y \neq 0$, $n_x = n_z = 0$)
  - z-axial modes, parallel to the z-axis ($n_z \neq 0$, $n_x = n_y = 0$)

- Tangential modes

  - y,z-tangential modes, parallel to the x-plane ($n_x = 0$, $n_y \neq 0, n_z \neq 0$)

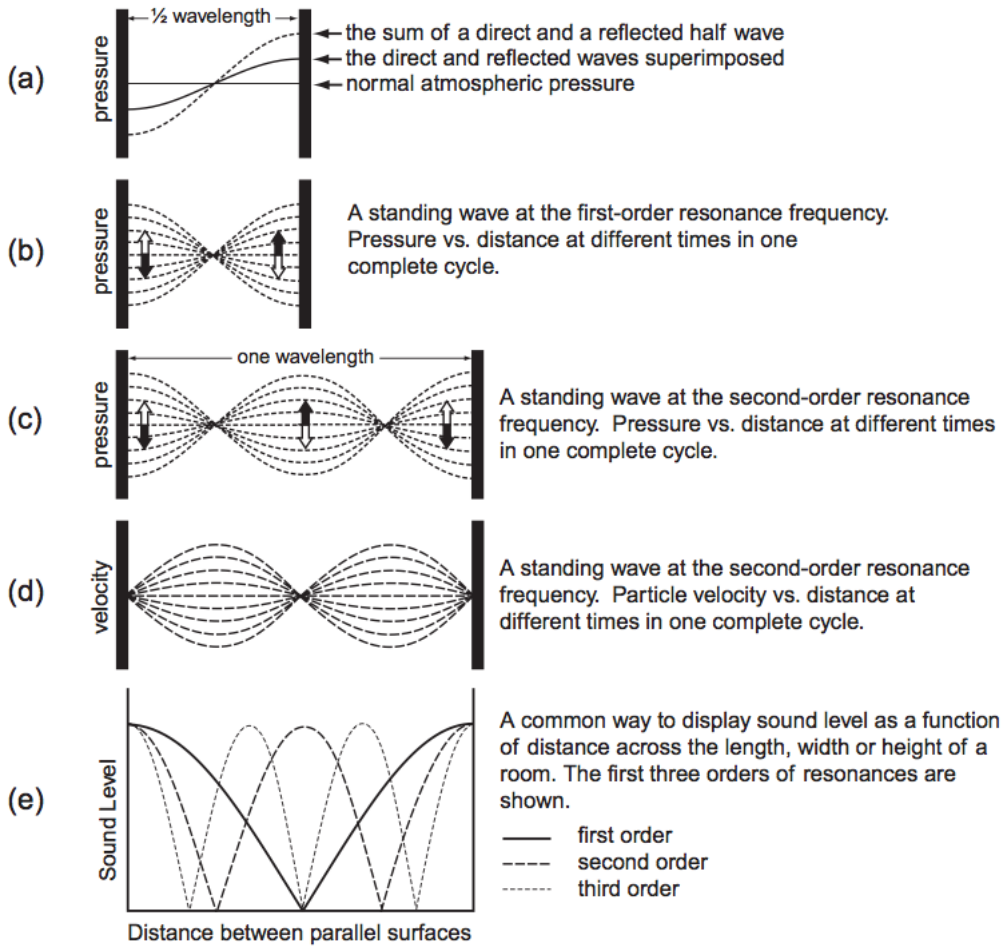


Figure 2.1: Representation of pressure and velocity in room modes ([50])

  - x,y-tangential modes, parallel to the z-plane ($n_z = 0,\ n_x \neq 0, n_y \neq 0$)
  - x,z-tangential modes, parallel to the y-plane ($n_y = 0,\ n_x \neq 0, n_z \neq 0$)

- Oblique modes ($n_x \neq 0, n_y \neq 0, n_z \neq 0$)

2.2.1 Effect of room dimension ratio on the frequency response and perception of modes

Room modes exist in all rooms, although some particular room shapes and ratios emphasize their effects. From the result obtained in the previous paragraph (2.15), it is understandable that the modal pattern depends on the shape of the room and its relative dimensions rather than the value of the dimensions themselves. This means that the position of room modes in rooms with the same aspect ratios should look very similar, besides the frequency axis scaling. In fact, larger rooms will have modes shifted downwards in the frequency axis. If the room is big enough, the lowest modes can be below the human hearing range, limiting their impact on the listening experience. On the other hand, small rooms will feature resonant frequency which are well above the human hearing lower limit, heavily modifying the perception of sounds in the room. This is shown in fig. 2.3, where the lower plot shows the position of axial room modes calculated with the aforementioned model (2.15) for a

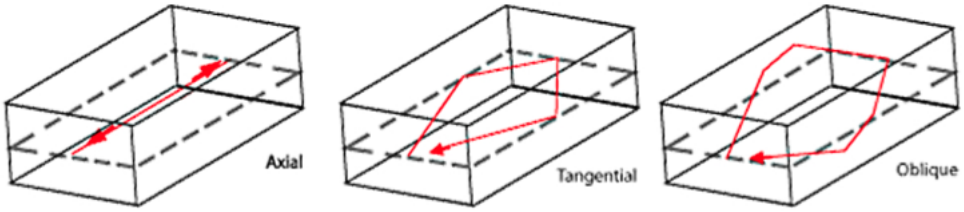


Figure 2.2: Axial, Tangential, Oblique room modes

room with dimensions that are exactly two times bigger than the one in the upper plot, with the same room ratios. It is clear that the relative position of the modes looks the same, but the frequency axis is different.

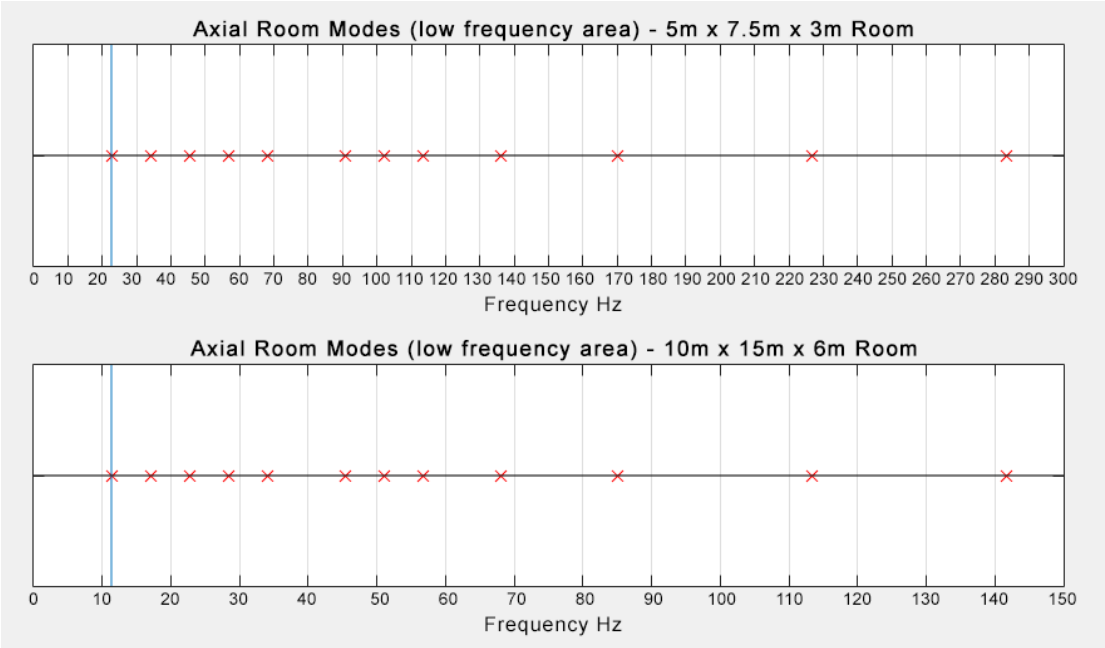


Figure 2.3: Axial modes in two rooms with different dimensions and same dimensions ratio

Also, in a room with two or three equal dimensions, modes relative to those surfaces sum their effect with each other, producing widely spaced resonances and adding even more energy to the few resonances relative to that dimension. These rooms are particularly problematic to treat because their frequency response's peaks have lots of energy, and are widely spaced. The reader is reminded that even in rooms where the ratio is different (also in non-simmetric rooms), room modes still exist.

2.3 Schroeder Frequency

As it is possible to understand from the previous paragraph, room modes occur for the lowest frequencies whose wavelengths are related to the room's dimensions. For higher frequencies this behavior does not happen, mainly because the wavelengths become small with respect to the small objects present in the room (such as the furniture) and geometrical and acoustic irregularities make it impossible to create and support resonances. The lower frequency range is dominated by separated resonances. In this context, the sound field can not be considered diffuse and the reverberation time concept is invalid, while a modal decay analysis, that focuses on the decay time of single resonances, is more appropriate. At higher frequencies, instead, the sound field can be considered as resulting from the constructive and destructive intereferences of many small acoustic elements.

Between these two regions, a transition region is present, where there is a gradual change between the two behaviors. At the centre of this region, the so-called Schroeder Frequency (or critical frequency) has been chosen as the frequency that divides the modal region from the statistical region. The Schroeder Frequency is defined as:

$$f_c \approx 2000\sqrt{\frac{T}{V}} \tag{2.17}$$

where V is the volume of the room in cube meters, and T is the reverberation time in seconds. The multiplicative constant has been changed from 4000 to 2000 after later studies ([46]). This means that the critical frequency is not only dependant on the dimensions of the room, but also on its furniture and acoustic treatment, which alter the decay time.

Below the Schroeder frequency, in the so-called "modal region", the frequency response is shaped by the presence of resonances caused by room modes. Fig. 2.4 ([4]) shows how the frequency response is dominated, at lower frequencies, by the presence of room modes that can be directly related to the specific dimension that generated them, and how they generally become more dense at higher frequencies, creating a flatter frequency response that allows the sound field to be considered statistically above the Schroeder frequency.

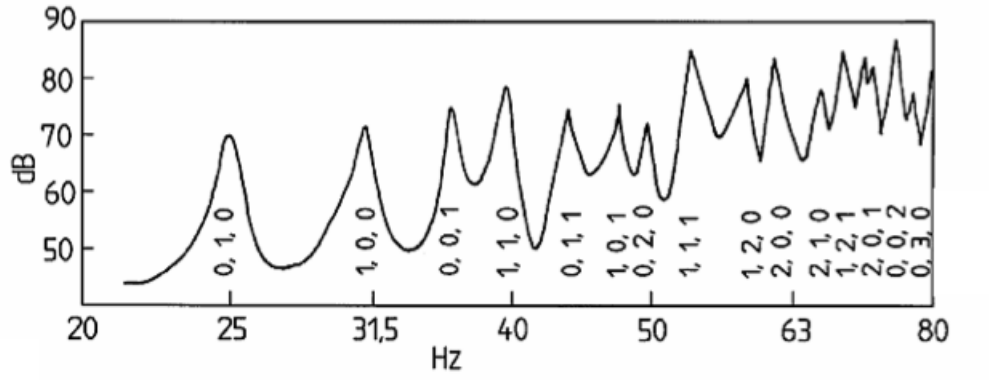


Figure 2.4: Density of room modes at very low frequencies and at increasing values of frequency ([4])

Fig 2.5 also schematizes this behavior.

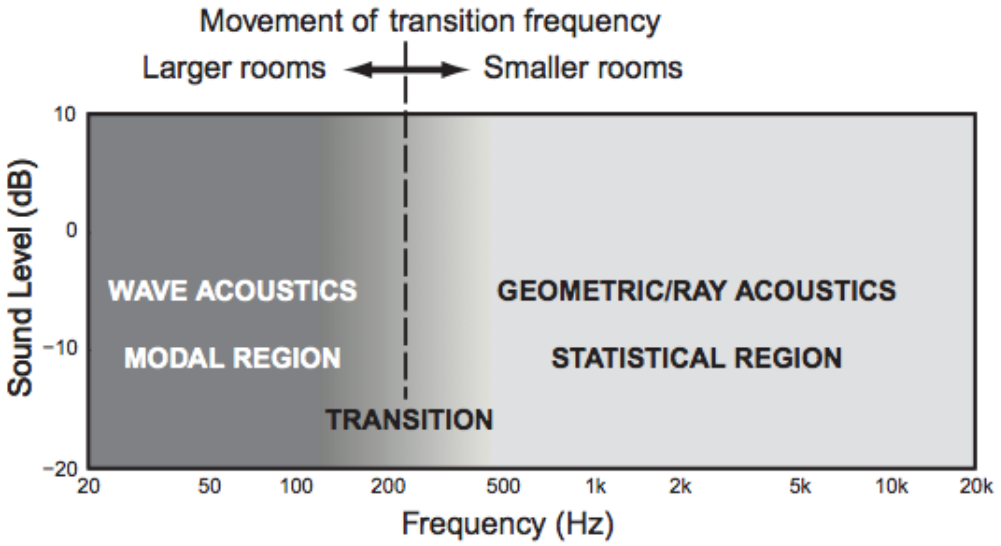


Figure 2.5: Frequency range areas divided by the Schroeder Frequency ([50])

Of course, modal activity is undesirable as the frequency response of the room becomes heavily dependent on the spatial localization of the source (speakers) and receiver (microphone, or the position of the listener's ears). Ideally, it would be better to have large rooms so that the Schroeder Frequency is moved down in the frequency range, possibly below the limits of the human hearing range, in order to avoid perceiving the effect of single resonances, which alter the perceived volume of selected frequencies. However, if the room is large enough to push the Schroeder frequency so low, it is possible that long reverberation time will be present, if the acoustic treatment stays the same while the room size gets bigger.

The time-domain counterpart of Schroeder frequency, called Mixing Time, is the time instant that separates the first part of a sound, composed of discrete arrivals (early reflections) from the statistical part (late reverberation). In [23], G. Ghelfi et al. studied the way in which the transition between the early reflections deterministic section and the stochastic diffuse field occurs, proving that the theoretical value of mixing time constitutes a valid approximation of the measured value only for rooms known as "sabinian", while they differ for non-Sabinian environments. Both terms will be explained in the next section.

2.4 Sabinian and Non-Sabinian Environments

The reader is reminded that a room can be considered "large" with respect to a certain frequency when its dimensions are many times greater than the wavelenght of that frequencies. Therefore, "small" rooms are such only for low frequencies, and can be considered large room above the Schroeder Frequency, where the statistical approach can be used. When a room is "large", it is assumed that the sound field is diffuse, which means that it is homogeneous (roughly equal in every point of the space) and isotropic (with sound arriving from all directions).

Jacobsen ([5] defines the concept of diffuse field with the following equivalent definitions:

- In a diffuse sound field, there is equal probability of energy flow in all directions.

  The statistical parameters characterizing a diffuse sound field are spatially homogeneous and isotropic.

- A diffuse sound field comprises an infinite number of plane propagating waves with random phase relations, arriving from uniformly distributed directions.

"Diffuse Sound Field" is actually just a theoretical concept, but under particular assumptions, this condition can be approximated with an high degree of accuracy. If a room features a sound field that can be considered, to a high degree of approximation, diffuse and isotropic, it is called "Sabinian" room, because Sabines assumptions regarding the soundfield are well approximated. Conversely, they are called "non-Sabinian" rooms if those conditions are not met. Some examples of Sabinian rooms are concert halls, auditoria, theaters, while non-Sabinian rooms are small rooms such as bedrooms, small control rooms, small studios (hence the reason why this classification is important to this research: all the rooms analyzed in this work are non-Sabinian).

In order to achieve a sound field that can be considered diffuse, some guidelines are given in [33] and [5] (which the reader is invited to read for more complete concepts regarding sound fields and reverberation time in closed environments), and can be summarized as follows:

- The shape of the room should not be extreme and have dimensions that differ radically from the others
- There should not be any focusing effects in the environment
- The irregularities of rough walls and presence of furniture and decorations can help scattering the incident sound energy
- The absorption coefficient should not be too extreme and the pattern of absorption in the room should not impede the formation of diffuse sound
- The room should not be too heavily damped

Kuttruff ([33]) stated that diffusely reflecting room walls alone do not guarantee that a sound field is diffuse. Of equal importance is the amount and distribution of wall absorption, because sound field diffusion is possible only if the absorption coefficient of the boundary is very small everywhere.

What is important for this research is the concept that, at low frequencies, the room behavior is defined almost exclusively by room modes, which have to be considered one by one since the sound field can not be considered diffuse.

2.5 Frequency Response

The Frequency Response characterizes the transmission dynamics of a system. If an input (as the simplest example, a sine wave with a given frequency) is injected into a system, a linear system will output another sine wave with a different magnitude and a phase modification. In order to account for both these effects, the Frequency Response is composed of the Magnitude Response and the Phase Response. The estimation of the frequency response is done by exciting the system with a scope signal which spans the range of the frequencies of interest (which, for audio applications, is the human hearing range) and measuring the output. In acoustics, an impulse (more precisely, a close approximation) is used, as impulses have very wide spectra. The input and output signals are compared to find the relation between them, either in the temporal domain (generating the impulse response) or in the frequency domain (generating directly the Frequency response). As far as room acoustics goes, the overall frequency response is actually a transfer function that takes into account the Loudspeaker's frequency response (which initially modifies the sound they emit), the Room's frequency response and the receiver's frequency response. For this reason, a room's frequency response is highly dependant on the position of both source and receiver, and this should always be kept in mind: the frequency response could be improved initially by simply moving these elements in the most appropriate point of the room. In fact, positioning the source and receiver on positions such as pressure nodes or antinodes can vastly change their ability of emitting sounds at resonant frequencies, or to perceive them. For more information on this, the reader is invited to read [50].

2.5.1 Magnitude Response

The Magnitude (or Amplitude) Response defines how the magnitude of the input signal is modified by passing through the system, by plotting the frequency range of interest on the x-axis ant a quantity in dB on the y-axis that characterizes the amplitude change of each frequency when passing through the system. In particular, the magnitude is the ratio between the ouput signal and the input signal at that frequency.

The Magnitude Response can be derived by the impulse response by applying a fast fourier transform and keeping the absolute value. It is necessary to keep in mind that the magnitude response describes the steady state response of the system, as if the system was excited with infinitely long sine bursts at each frequency and the output was measured when each frequency has reached their final, steady state value. Fig. 2.6 shows one of the impulse responses used for this research, measured by the Suono e Vita staff, and the magnitude response (in the 20-300 Hz area) generated by applying, in Matlab™, the absolute value to the fast fourier transform of the impulse response itself.

In the following paragraphs, the term "Frequency Response" will refer to the Magnitude Response, in accordance to the general meaning used in the scientific community.

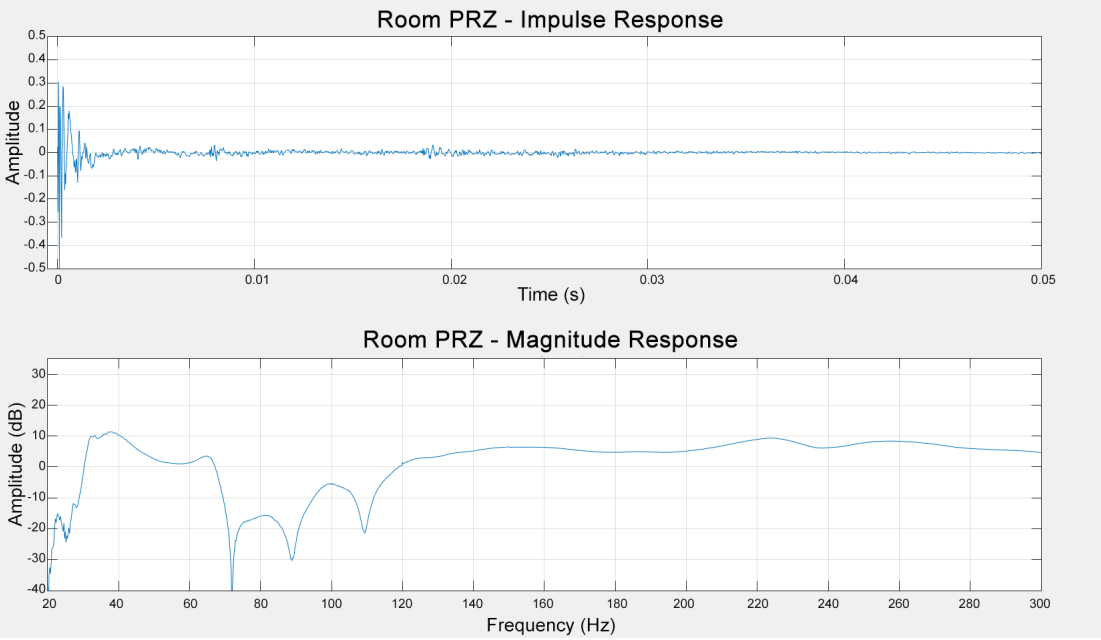


Figure 2.6: Impulse response and Magnitude response of Room PRZ

2.5.2 Phase Response

The phase response describes how much different frequencies will experience a phase shift when passing through the system. In particular, the phase response is measured in radians or degrees (y-axis) with respect to frequency in hertz (x-axis), and is defined as the difference between the phase of the output signal and the phase of the input signal (which is set by definition to zero). The y-axis is not limited to the range between 0 and 360 degrees, as the phase shift can accumulate above the higher limit. The derivative of the Phase Response with respect to frequency is called Group Delay and it describes the temporal amount by wich the amplitude of each frequency component is delayed passing through the system. Of course, group delay should be as small as possible in order not to be audible. In ([31]), Blauert and Laws defined a table with thresholds of audibility for the group delay at different frequencies.

It is important to focus on the fact that transient sounds exhibit phase slopes, representing the rate of change of phase with frequency. A transient contains a great number of frequencies (delta functions contain all frequencies), so each frequency will have its own individual phase shift. The superposition of differently shifted sine waves can create greatly varying signals at the end. Fig. 2.7 ([39]) shows how this happens: the sum of the same frequencies, but shifted differently, creates a total sound which varies considerably ([39]).

Phase response is quite problematic to work with, since any equalization applied to the system in an attempt to equalize the magnitude response, also has an impact on the phase response, therefore on the delay of different frequency components. The importance of treating both the magnitude and phase response has been discussed in [26]. It is important to remind the reader that this thesis work started from a phenomenologic analysis of impulse responses of real rooms, and the starting point was [37]. Since the author wanted to inspect if the magnitude response was always significative in the perception of sound levels also for short sounds, the phase

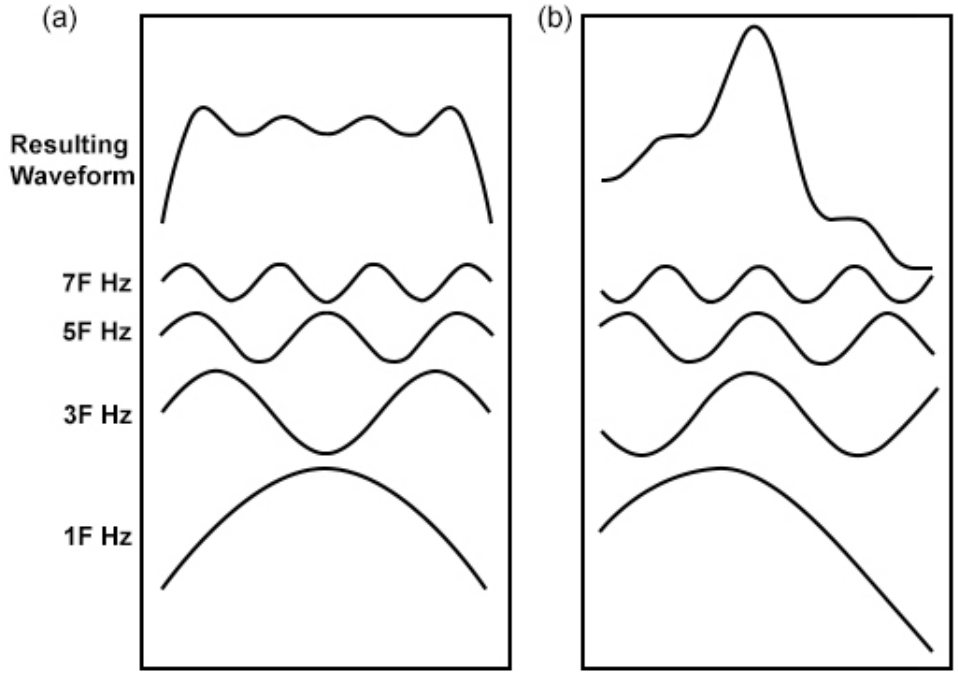


Figure 2.7: Phase Slope effects on a transient sound ([39])

response was not analyzed in this research. Further developments should take into account also the room's phase response and analyze its impact on the results of this work.

2.6 Resonances, Q Factor and Modal Decay

Undamped resonances will resonate strongly in time and will be clearly visible in the spectrum as sharp peaks. Instead, damped resonances will have lower peaks and tend to blend with neighboring ones ([35]). The Q factor in the frequency domain is defined as the ratio between the resonant frequency and its -3 dB bandwidth:

$$Q = \frac{f_o}{\Delta f} \tag{2.18}$$

At low frequencies, where the room behavior is non-Sabinian, the concept of Reverberation time loses its meaning and a specific decay analysis for each room mode has to be carried out instead. The decay time of each resonant frequency is closely related to the room damping at that frequency. The higher the Q factor (narrow frequency peaks, low damping) the longer is the decay time, and it will be mostly related to that single room mode. Conversely, when the Q factor is low, the decay time is shorter and influenced by the combined effect of close frequencies ([35]). As it happens, the fact that higher and narrower peaks tend to have longer decay times is well known and can be seen by any waterfall plot, such as fig. 2.8 wich shows the waterfall plot of Room CN, one of the rooms used in this research work.

As stated in [14], both the modal Q factor and hence the temporal behavior of the stimulus are important in the detection of isolated resonances. This A lower

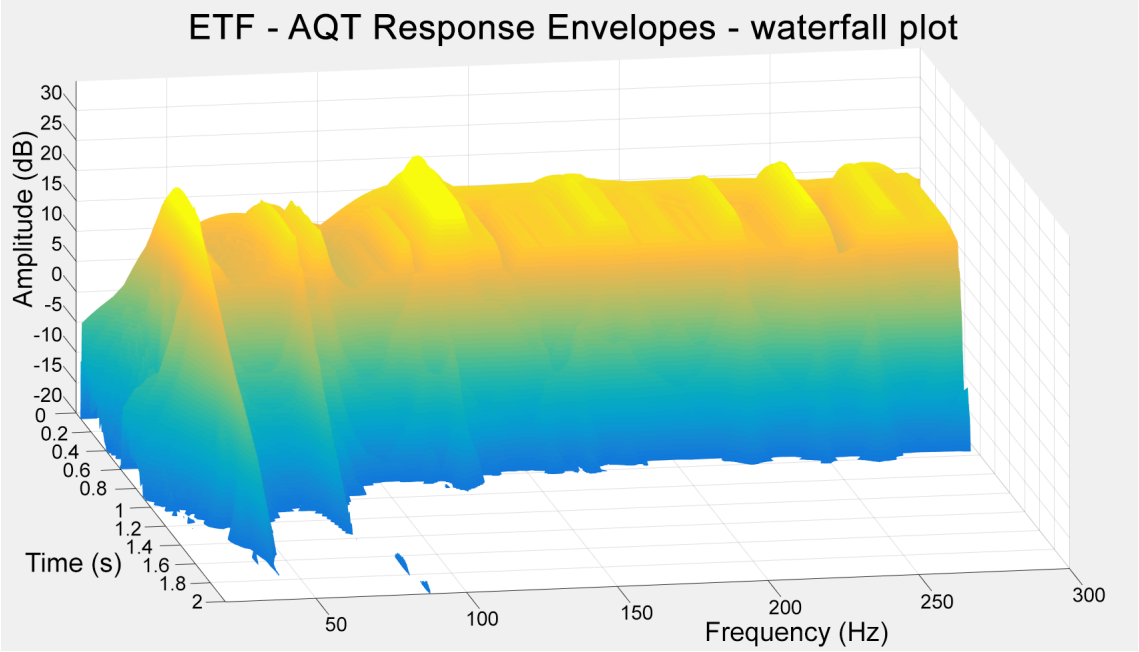


Figure 2.8: Waterfall plot of Room CN - longer modal decay times on frequency response's sharper peaks

threshold of 16 has been defined for the subjective perceptibility of the Q Factor in [14], meaning that a Q Factor of at least 16 is required to detect a presence of a resonance. This value suggests the threshold below which further room treatment would be unnecessary, but it also decreases considerably with increasing frequency.

A common interpretation of the 60 dB modal decay time, relating the Q factor with the center frequency of the resonance, is given by eq. 2.19 ([32]) and is measured in seconds:

$$T = \frac{2.2Q}{f_0} \tag{2.19}$$

The Q factor of a mode can also be calculated if the decay time on a $e^{2\pi}$ decay from the steady state value if it is available ([35]).

Beranek ([34]) provides a description of what happens when a sound source is turned on in a small enclosure. If the source is assumed to be constant in strenght and containing a single frequency which is one of the normal frequencies of the enclosure, the average (in time and space over a wavelength) rms pressure level will build up (a large portion of our research focuses on how exactly this buildup occurs) until it reaches the value:

$$|p_n| = \frac{K}{k_n} \tag{2.20}$$

where K is a constant dependant on the position and strength of the source and the volume of the room, and k_n is the room damping constant, depending on the room's dimensions and absorption. If the frequency does not coincide with one of the normal frequencies, the final rms pressure level will be in accordance to the curve in fig. 2.9 ([34]). If the driving frequency lies between two normal frequencies, or if k_n is large so that the resonance curve is broad, more than one normal node

of vibration will be excited significantly. The larger the room, and the higher the frequency, the nearer will be the normal frequencies and the more of them will be excited by one frequency.

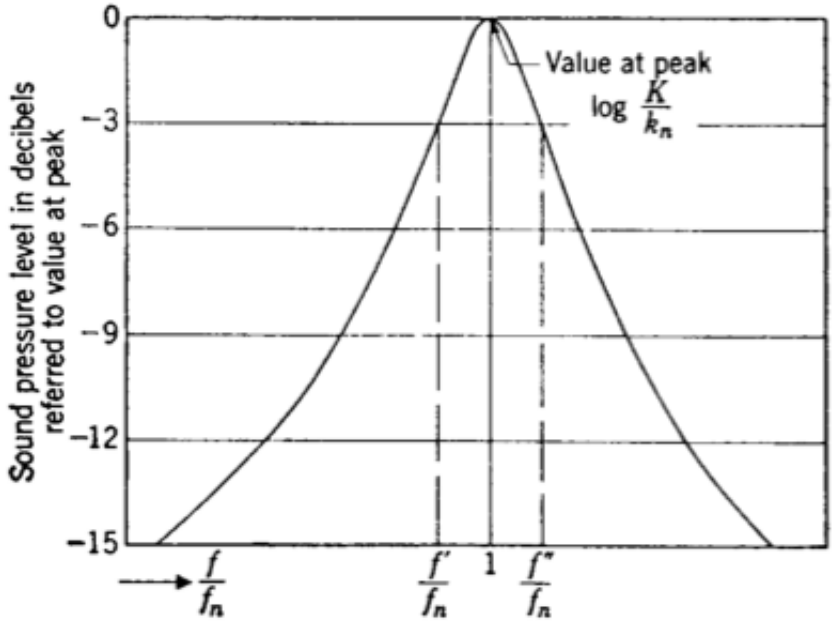


Figure 2.9: Resonance curve for a normal mode of vibration - sound pressure level vs. the ratio of frequency to f_n ([34])

When the sound source is turned off, each normal mode of vibration behaves like an electrical parallel-resonance circuit in which energy has been stored initially. The pressure for each normal mode of vibration will decay exponentially at its own normal frequency, as shown in fig. 2.10 ([34]). If only one mode of vibration was excited (Fig. 2.10, subfigure a), the decay is given by:

$$|p_n| = \frac{K}{k_n} e^{-k_n t} cos(\omega_n t) \tag{2.21}$$

If two or more modes of vibration were excited, beats occur and, if they have different decay constants, the shape of the decay will be composed of two different slopes.

It is important to highlight that the damping constant determines the maximum height and width of the steady-state resonance curve, as well as the rate of the decay after the sound source has been turned off. The room acts like an assemblage of resonators that act independently of each other when the sound source is turned off. Further research should be made in this direction, in light of the results of this research work that hint at a strong relation between small environments and higher order systems in regards to their underdamped and overdamped behavior.

2.7 Reverberation Time

The reverberation time is one of the most important and immediate acoustic parameters, as it gives a general idea of the acoustic behavior of the room. It is defined

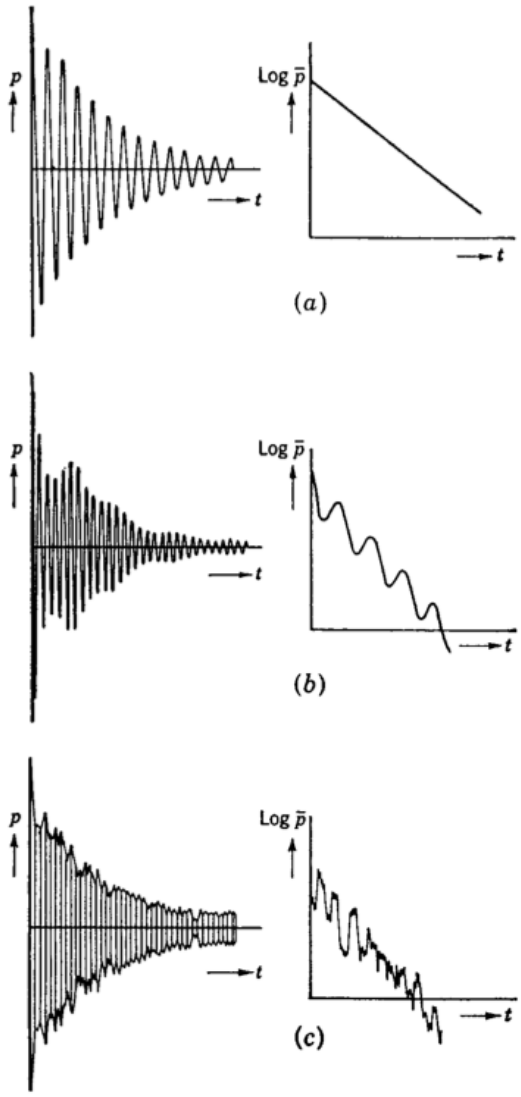


Figure 2.10: Sound pressure decay curve for a single mode of vibration (a), for two closely spaced modes of vibration with the same decay constant (b), and for a higher number of closely spaced modes of vibration with the same decay constant (c) ([34])

as the time (in seconds) necessary for the sound field to decay by 60 dB from its steady state level by Sabine's equation:

$$T_{60} = 0.161 \cdot \frac{V}{\sum_{i=1}^{n} \alpha_i S_i} \tag{2.22}$$

where V is the volume of the room in cube meters, α_i is the absorption coefficient of the i-th surface and S_i is the area of the i-th surface. The member at the denominator can also be called Equivalent Absorption Area of the room.

2.7.1 T_{30}, T_{20}, Early Decay Time EDT

Similar metrics exists, called T_{30} and T_{20}, which are used when the noise floor is too high to correctly record a fall of 60 dB in the sound field. T_{30} is defined as the time needed for the sound field to decay by 30 dB, multiplied by 2; T_{20} is defined as

the time needed for the sound field to decay by 20 dB, multiplied by 3. The Early Decay time, similarly, is defined as the time needed for the sound field to decay by 20 dB, multiplied by 6. This index arises from the necessity of differentiating the initial and final slopes of the decay, as it has been observed that the first part can have a different slope due to the directivity of the source and the shape, position and reflectivity of nearby objects.

All versions of reverberation time can be computed more easily and starting from the room impulse response using the Schroeder Backward Integration method.

2.7.2 Schroeder Backward Integration method

Introduced by Schroeder ([45]), this method consists of a reverse time integration that turns the squared impulse response into a decay curve. The decay as a function of time can be written as:

$$E(t) = \int_t^\infty p^2(\tau)d\tau \quad = \quad \int_\infty^t p^2(\tau)d(-\tau) \tag{2.23}$$

where E is the energy of the decay curve as a function of time, t is time and $p(t)$ is the sound pressure of the impulse response as a function of time. For the integrated impulse response method, the steady state level is the total level of the integrated impulse response. Norm ISO 3382 describes the way backward integration is performed. A least-squares fit line is computed for the curve (or fitted manually). The slope of the straight line is the decay rate, d, in decibels per second. Reverberation time is then computed as:

$$T_{60} = \frac{60}{d} \tag{2.24}$$

The same formula as 2.24 can be used for calculating $T30$ and $T20$, but the range over which the slope is calculated is respectively 30 and 20 dB. Specifically, the evaluation range is from 5 dB to 35 dB below the steady state level for T_{30} and from 5 dB to 25 dB below the steady state level for T_{20}. In all cases, for a maximum underestimation of 5 percent, the level of the background noise must be at least the evaluation range plus 15 dB below the maximum of the impulse response. As an example, for the measurement of T_{30}, the level of the background noise must be at least 45 dB below the maximum.

These values can be computed for octave bands or third of octaves. In this research, the backward integration has been performed on the response envelope of each convolved pure tone burst (see chapter 5) in order to find the decay time of each frequency. However, for frequencies with a very low steady state value (such as frequency response's valleys) the noise floor was very close to the steady state level, producing unnaturally long results in the decay time. This has lead to the definition of a new parameter that solves this problem.

Chapter 3

Psychoacoustics and Subjective Audibility

In order to correctly understand the reasoning behind the developmente of the two psychoacoustic tests on which this research is built on, it is necessary to introduce the basic concepts of psychoacoustics, that describe the relationship between the properties of sound and the subjective perception of those properties by listeners (and their lateral effects). If the reader is interested in learning more about psychoacoustics, two suggested reads are [12] and [38].

Psychoacoustics is the scientific study of sound perception. More precisely, it is the branch of science that studies the psychological and physiological responses associated with sound ([12]). As a matter of fact, the perception of sound is not only related to the properties of the source; after traveling through the air, sound waves are perceived by our ear and transformed into electrical pulses that are then interpreted by our brain. For this reason, any malfunctioning or natural effect that impacts the ability of the ear to capture sound (such as the progressive loss of hearing ability in the high frequency range that happens naturally with aging), will impact the perception of the sound itself, leading to a different listening experience. The anatomy of the hearing system, which is well known, is able to explain most of the principles that alter the psychoacoustic perception of sound, but it fails to explain the most complex behaviors ([38]). For this reason, psychoacoustic testing is still an important research method: by controlling the variables of the test sounds, it is possible to understand how listeners react and what do they perceive.

Psychoacoustics have lots of application fields including software development, digital signal processing, audio production and many more: as an example, its principles are used consistently in hearing aid design (alleviation of hearing loss) and cochlear implants. Evaluation of room acoustics and room design also are heavily influenced by the limen of perception of many variables such as presence of resonance modes, decay times, reverberation time. If the values of such variables fall under the threshold that humans can detect, there is no need to furtherly treat the room. Perceptual audio coding techniques also rely heavily on these principles, such as the mp3 (and, more generally, all audio dynamic range compression techniques) which discards the frequency and temporal information which are found to be not perceivable by the ear, avoiding to encode frequencies which are under the perceivable threshold (like ultra high and low frequencies, and those who are temporarily masked by other, louder frequencies) and reducing considerably the files'

dimensions. As another example, some speakers are built to emphasize the low frequency perceived emission even though their physical dimension would not allow them to emit ultra low frequencies. Psychoacoustic principles are also used in the music industry and music production. Some plugins used in audio production allow to increase the upper harmonics of low frequency instruments, making it possible to perceive an improved bass response even on small earphones by exploiting the principle of the "missing fundamental effect" (the ability of the ear to reconstruct the ultra low fundamental frequency when overtones are present). Since the hearing threshold is different for all frequencies (see Fletcher-Munson curves), it is possible to understand that, increasing the spectral content at frequencies with a lower hearing threshold, the sound will be perceived as louder. This has also been one of the causes (alongside the dynamic range compression) of the so-called "Loudness war", the tendency to over-compress music in order to raise the perceived volume.

This research focuses on the perception of levels of non-repeated, separated sounds and the perceived precision loss when a sound is played in a room, using headphones. Therefore, mainly loudness, pitch perception and resonances perception concept will be recalled. For basic concepts such as the anatomy of the human ear, and more advanced concepts whose principles did not play a role in this research work, such as masking, space perception, speech perception, pattern perception and more, the reader is invited to read [38].

3.1 Loudness Perception

Loudness is defined as that attribute of auditory sensation in terms of which sounds can be ordered on a scaling extending from quiet to loud ([38]). The human ear has a very large dynamic range, meaning that the ratio between the loudest sound perceivable without directly damaging our ears and the quietest sound that can be perceived is huge, in the range of 10^{12} to 1 (about 120 dB).

3.1.1 Hearing Threshold

The hearing threshold with respect to frequency can be measured in two ways: MAP (minimum audible pressure) is measured by a small probe microphone as closely positioned as possible to the eardrum, and usually using headphones. MAF (minimum audible field) instead uses speakers and calculates this parameter after the subject left the environment, positioning a microphone where the tester's head was located. Both parameters are measured for all frequencies and have a similar behavior, however they show some differences caused mainly by the effect of the shape of the head, torso and pinna. A different behavior happens at medium frequencies, caused mainly by a resonance produced by the meatus and pinna. For both curves, shown in fig. 3.1 ([38]) the hearing threshold is lower at mid frequencies, whereas it is mich higher at very low and very high frequencies, meaning that, in order to be perceived, a low frequency sound has to possess high sound pressure levels, contrary to mid frequencies sounds. This is caused, at least partly, by the tramsission characteristics of the middle ear.

Regarding the frequency range, the human auditory system is said to be able to perceive frequencies between 20 Hz and 20 kHz, even though these values are not exact for all subjects. What is known, however, is that the auditory range changes

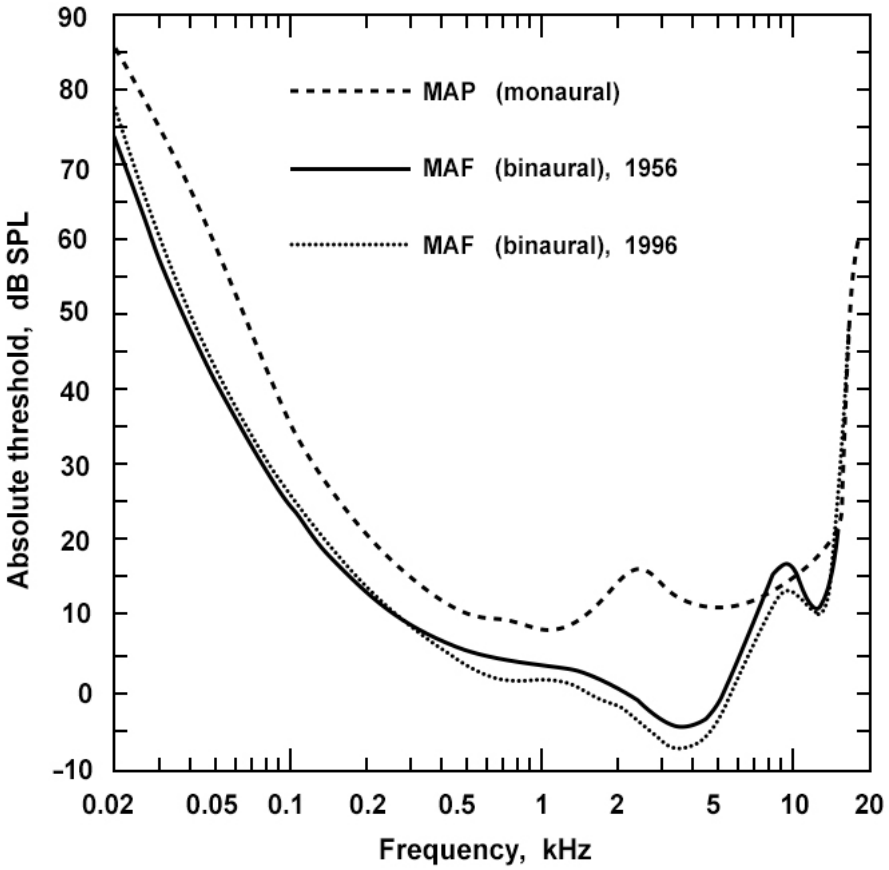


Figure 3.1: Minimum Audible Field and Minimum Audible Pressure curves with respect to frequency ([38])

with age, reducing the capacity of the subject of perceiving ultra low and ultra high frequencies. This loss of sensitivity is much more pronounced for higher frequencies.

3.1.2 Equal Loudness Curves

Equal Loudness Curves (also known as Fletcher and Munson curves) where initially defined by Fletcher and Munson in 1993 at Bell Labs. These curves can be created by letting testers hear a 1 khz pure tone at a set playback level, and other tones of which the tester has to adjust the volume to match the one of the first sound. This is repeated for many frequencies, and for different listening levels. These curves have been measured in different laboratories and context with slightly different results, even though the general behavior is always the same. Therefore, the exact shape of these curves should be taken with caution. In particular, they have been measured again in 1956 by Robinson and Dadson and standardized in 1986. The standard was later corrected in 2003 by averaging the results collected by twelve international studies ([12]).

Equal loudness curves, shown in fig. 3.2, attempt to describe the sound pressure level that different frequencies have to possess in order to be perceived at the same level, for different loudness levels.

It is clear, therefore, that the sensitivity of the human ear is not the same at all frequencies, which means that sound waves with the same pressure level can indeed

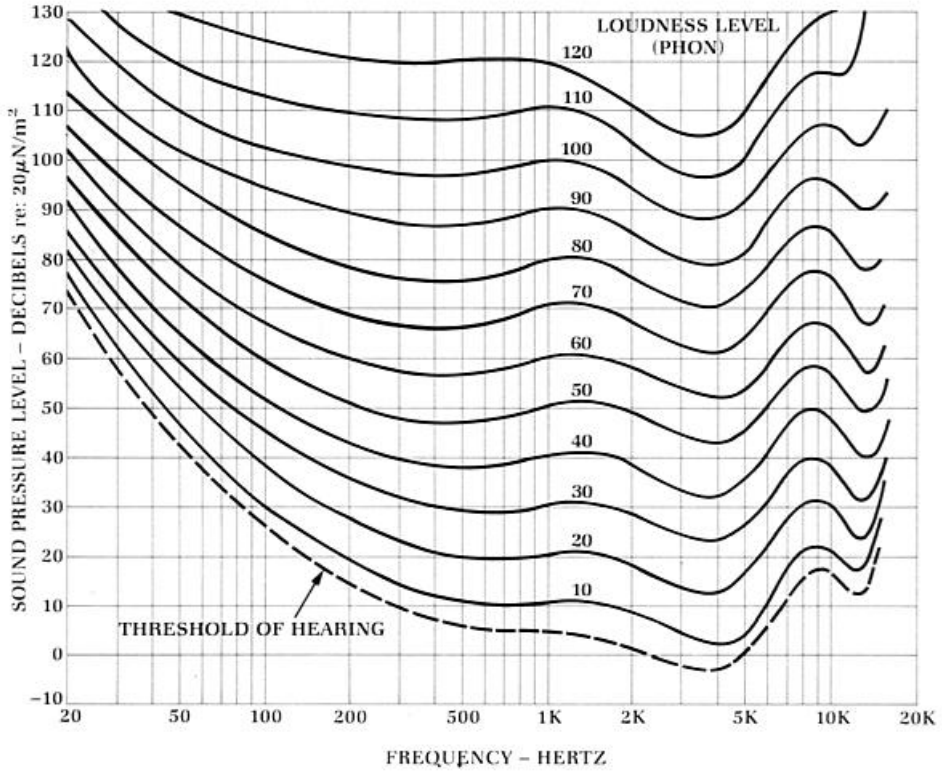


Figure 3.2: Equal Loudness Curves (Fletcher-Munson curves)

be perceived as having very different loudness, depending on their spectral content. Interestingly, the curves become flatter if the playback level is louder, meaning that at higher volumes the hearing threshold of low and high frequencies is closer to the middle frequencies one, therefore making such sounds more perceivable. This is what lead to the impression that "louder sounds better", also a key concept in what lead to the Loudness War.

The shape of Equal Loudness curves has been often used in the design of sound level meters attempting to measuring the loudness of complex sounds. Such meters incorporate weighting curves in order to weight differently every frequency according to the curves: the "A" weighting (based on the 30 phon equal loudness contour) is used at low sound levels, diminishing the contribution of low frequencies; the "C" weighting (based on the 100 phon curve), which is flatter and introduces less modifications, is used at high playback levels, while the "B" weighting (based on the 70 phon curve) is used at intermediate levels.

Moore ([38]) already introduced the concept that these curves can be reliable only with steady state sounds of relatively long durations, while "the response to transient sounds does not correspond to the subjective impressions of the loudness of such sounds". It is important to remember also that the loudness perception of a sound depends on its spectral spread, because complex sounds have a large spectrum, and the ear divides the incoming sound in spectral bands related to the concept of critical bands and this can lead to differences in the loudness perception.

It is known that both the loudness of sounds and the absolute thresholds depend upon the sound duration. Moore ([38]) states that, for long durations exceeding about 500 milliseconds, the sound intensity at threshold is roughly independent of

duration, while for durations shorter than about 200 milliseconds, the sound intensity necessary for detection increases as the duration decreases. This is very important and closely related to this research, because it can be related to sounds that reach their actual steady state values, and sounds which do not. For this research, tests were conducted with 150, 250 and 550 milliseconds bursts (see Chapter 5).

Regarding this research, many test questions compared sounds asking questions regarding their perceived loudness. All sounds had at least a few Hz of difference, therefore were of course subject to the equal loudness curves. The questions were developed in order to understand if the problems in the room's frequency response were strong enough to alter the perception of loudness with respect to the one predicted by the equal loudness curves, and to what extent they could have an impact on transient sounds.

3.1.3 Detection of Intensity Changes and Weber's Law

The issue of measuring the smallest detectable change in intensity lead to the adoption of different methods that compare different types of stimuli. Normally, a two-alternative forced choice (2AFC) method is used, in which two stimuli with different levels of the variable under test (in this case, intensity level) are presented to the listener in random order. The threshold of detection of intensity change is usually defined as the value which produces 75 percent correct responses ([38]). Three slightly different methods for measuring intensity discrimination exists (modulation detection, increment detection, intensity discrimination of gated or pulsed stimuli), even though they produce results with the same general trend. For the method "Intensity discrimination of gated or pulsed stimuli", in which the listener has to indicate which one is louder after comparing two pulses of sounds, the change in level at threshold of audibility is ([38]):

$$\Delta L = 10 \log_{10}((I + \Delta I)/I) \tag{3.1}$$

where I is the intensity of one sound, $(I + \Delta I)$ is the intensity of the other, meaning that the smallest detectable intensity change is approximately a constant fraction of the intensity of the stimulus (that is, $\Delta I/I$ is roughly a constant and is called "Weber constant"). This is an example of Weber's Law, which states that the smallest detectable change is proportional to the magnitude of the stimulus.

The behavior has been observed to be slightly different for pure tones and noise, for which instead of being a constant, the Weber fraction actually decreases when the level of the initial sound (I) is increased. This is called "Weber's Law near miss", and according to this behavior, it seems that the smallest detectable intensity actually improves at higher volume. This fact contradicts the expected behavior of the auditory system, that should cause the Weber's ratio to increase (therefore deteriorating the discrimination) with higher sine tone levels.

For this research, the author performed selected questions of the first psychoacoustic test at two different listening levels: one (quieter) which was the same as the rest of the questions, and another, louder. Results encouraged to keep on utilizing the "quieter" playback level also for the second psychoacoustic test, as most testers were able to answer questions with more confidence and, generally, expected results.

This is also expected, since none of these questions (and almost none of the second test) featured pure tones.

3.1.4 Loudness Adaptation and Fatigue

In human sensory systems, the exposure to a stimulus of great duration and intensity produces changes in the system's responsiveness. Such changes may appear during the presentation of the stimulus (changing the perceived apparent magnitude of new sounds with respect to the one that would be perceived if the loud stimulus was not present), or after the stimulus has ended (for example, shifting the auditory threshold of certain frequencies). Since this research work deals with psychoacoustic tests performed on headphones, it is important to design such tests in a way that listeners are not fatigued by the sounds, which would modify their answers from the ideal behavior.

For a complete explanation of such phenomena, the reader is invited to see [15].

Auditory Fatigue

Fatigue results from "the application of a stimulus which is usually considerably in eccess of that required to sustain the normal physiological response of the receptor, and it is measured after the stimulus has been removed" ([29]). This is often referred to as "post-stimulatory auditory fatigue" and the shift in threshold is called temporary threshold shift (TTS). It is measured by first measuring the subject's absolute threshold at a particular frequency, and then measuring it again after the subject has been subject to a fatiguing stimulus for a certain period of time. One problem with this approach is that the recovery can be quite fast, therefore the measurement has to be done quickly, which may, in turn, introduce inaccuracies. Some factors that influence the TTS are ([38]):

- The intensity of the fatiguing stimulus (I)
- The duration of the fatiguing stimulus (D)
- The frequency of the fatiguing (exposure) stimulus (F_e)
- The frequency of the test stimulus (F_t)
- The time between cessation of the fatiguing stimulus and the post-exposure threshold determination, called recovery interval (RI)

In general, TTS increases with I and this behavior depends on the intensity: at low levels, TTS changes slowly as a function of I and occurs for test tones with frequencies F_t close to F_e. For higher intensities (I), the frequency range over which this effect happens gets broader. When F_t is higher than F_e, TTS grows very rapidly as a function of I. For very high intensities, the TTS grows even more rapidly, and this is said to divide a physiological type of fatiguing from a more pathological and permanent one ([29]). Fatigue also increases with the exposure duration D, and has been found to be linearly related to $logD$, when D is greater than 5 minutes ([29]). Fatigue effects are generally more evident at mid and higher frequencies; when the fatiguing stimulus is broadband noise, the maximum TTS occurs between 4 and 6

kHz ([38]). Furthermore, TTS generally decreases with increasing RI. The recovery curve is diphasic, meaning that after an initial decrease in the fatigue, a small bump is observed before continuing to decrease.

Regarding music, the effect of listening to very loud (110, 120 dB) music, for example at rock concerts, is of course negative and may produce permanent hearing damage. Even though many long-time musicians do not show huge hearing degradation, hinting at the fact that music (or pleasant sound) may have an impact that is not as negative as thought [38], professionals as well as casual listeners should always wear earplugs when they are subject to high playback levels, especially for prolonged periods of time. In fact, the permanent damage caused by the exposure to intense sounds is related to the total amount of energy to which the ear is subject over a given period. Louder sounds can, of course, produce permanent damage very quickly.

Auditory Adaptation

Auditory adaptation can be measured with different methods. Some of them include:

- Applying a tone of a fixed level to one ear, and ask the subject to balance it to a tone of the same frequency, but with variable level, applied to the other ear. Remove the tone in the first ear, and repeat the loudness match after three minutes (adaptation period). Generally, the tone in the control ear should produce a loudness match at a lower level. This method is called Simultaneous dichotic loudness balance (SDLB).

- Asking the listener to match the loudness of tones with different frequencies, presenting the adapting tone continously to one ear and the comparison tone intermittently either to the same or the other ear

- Asking the listener to match the loudness of a continously presented sound as to maintain it at constant loudness (usually, the subject increases the volume slightly with time, indicating that adaptation is occurring)

- Asking the listener to assign numbers that rate the perceived loudness of a sound at sucessive time intervals (Scharf, 1983).

Different results arise from different methods, but some general consensus is that there is no significant loudness adaptation for adapting tones between 50 and 90 dB SPL ([38]); Scharf ([43]) stated that a sound, presented alone, adapts only if it is below 30 dB SPL; high frequency pure tones adapt more than low-frequency ones or than noise, and that steady state sounds adapt more than modulated sounds. It looks like there is no significant interaction between variables such as the person's threshold, sex, or age (even though children under 16 years old seem to adapt less than adults) and the degree of adaptation of the individual. As a matter of fact, the degree of adaptation varies widely from subject to subject: while most people hear a decrease in level over time as a result of adaptation, some people do not, and some actually report that the tone disappears.

With regards to this research, the psychoacoustic tests were designed in order to avoid fatigue and adaptation as much as possible. Sounds were very short (most of

them were under 550 milliseconds), reasonably quiet, monaural (avoiding the possibility of introducing adaptation in one of the years), with content heavily focused in the low frequency range, and with periods of silence between one question and another which were long (a few seconds) if compared to the test sounds.

3.2 Pitch Perception

A sound can be described, in the spectral domain, as the sum of its frequency components. Namely, the lowest frequency is called the fundamental frequency, and it is related to a certain wavelength and period wich are used to name the note. Other significant frequencies in the spectrum of a sound are the fundamental's overtones, which are composed by its harmonics (whole number multiplies of the fundamental frequency) and partials (other overtones which are not multiple of the fundamental frequency). Sometimes, the fundamental is not the dominant frequency; as an example, the dominant frequency for the transverse flute is double the fundamental frequency. If the sound is composed mainly of harmonics, its behavior in the time domain will be mostly periodic and the sound can be defined as "pitched", while, if the contribution of harmonics is lower than the contribution of other partials, the sound is not periodic and it is defined as "unpitched".

Pitch has been defined by American Standards Association, in 1960, as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale". Pitch is related to the period of the waveform of a sound, which corresponds to the frequency for a pure tone and to the fundamental frequency for a complex tone. However, the pitch itself is a subjective quantity, therefore it cannot be measured directly. Assigning a pitch value to a sound is generally understood to mean specifying the frequency of a pure tone having the same subjective pitch as the sound ([38]).

There are different theories that try to explain how the physical sound is translated by the auditory system to give the impression of pitch. In general, theories can be divided into:

- Place coding

  This theory is based on two distinct postulates:

  - The stimulus undergoes a spectral analysis in the inner ear, so that different frequency components excite different places along the basilar membrane and hence neurones with different characteristic frequencies. This postulate has been confirmed and proved true in a number of different ways.
  - The pitch of the stimulus is related to the pattern of excitation produced by that stimulus. This postulate is still matter of dispute.

- Temporal coding

  This theory suggests that the pitch of a stimulus is related to the time pattern of the neural impulses evoked by that stimulus.

Both theories do not completely explain the way pitch is perceived. In fact, Temporal coding theories would not work for frequencies above 5 kHz, while Place

coding theories do not explain phenomena such as the so called "missing fundamental" one. For a complete and thorough explanation of both theories, the reader is invited to see [38].

3.2.1 The phenomenon of the missing fundamental

It may happen, sometimes, that the fundamental frequency of a note is below the limit of the human hearing, or that a sound is synthesized or processed so that it possesses only the upper harmonics, but it is missing the fundamental frequency. In this cases, the brain is capable of reconstructing the fundamental frequency and gives the impression that the fundamental frequency is present, introducing possibly only a change in timbre. This is not at all uncommon, and even when the fundamental frequency is present, the pitch of the tone is usually determined by the harmonics and a low pitch can be heard when there is no component of that frequency. This behavior, due to the brain interpretating the repetition patterns, poses some doubts on the classical "place" theory that relates the perceived pitch with the position of maximum excitation on the basilar membrane when regarding complex tones.

It was once thought that the cause of this behavior was the introduction of distortions caused by the anatomy of the ear. However, experiments proved that this behavior still held true even after the addition of noise that would have masked such distortions if they had been present. The precise way in which the brain processes the information present in the overtones to calculate the fundamental frequency is still a matter of debate.

The pitch of the missing fundamental is not always perceived, however; it appears that, for narrow stimulus conditions with small number of harmonics, listeners can be divided between those who do perceive missing fundamentals, and those who do not ([6]).

As already introduced, the concept of the missing fundamental being reconstructed starting from the overtones has been exploited to create the illusion of bass in sounds systems which were not capable of emitting very low frequencies. As an example, products such as MaxxBass™ plugin from Waves Audio™ synthesize higher harmonics creating the impression of the presence of lower frequencies even in small playback systems.

3.3 Timbre Perception

Timbre has been defined by the American Standards Associations (1960) as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar". It is a quite complex feature to define, to the point that it has also been defined as "...the psychoacoustician's multidimensional waste-basket category for everything that cannot be labeled pitch or loudness." ([49]).

The perception of timbre has been studied through many psychoacoustic experiments in the past. Such experiments have demonstrated that one of the features that characterizes the most the perception of timbre is actually the onset (attack) of the sound; as a matter of fact, if a note is played on a violin open string and a trumpet get their onset and offset removed, it becomes quite difficult for listeners to tell them apart. Therefore, lots of characteristics of the instruments themselves

Subjective experience	Objective physical phenomena
Tonal character, usually pitched	Periodic sound
Noisy, with or without some tonal character, including rustle noise	Noise, including random pulses characterized by the rustle time (the mean interval between pulses)
Coloration	Spectral envelope
Beginning and ending	Physical rise and decay time
Coloration glide or formant glide	Change of spectral envelope
Microintonation	Small change (one up and down) in frequency
Vibrato	Frequency modulation
Tremolo	Amplitude modulation
Attack	Prefix
Final sound	Suffix

Table 3.1: Relation between subjective experiences and related physical phenomena, regarding timbre (Erickson)

are very relevant to the perception of the timbre, like the scraping of the bow on a stringet instrument, the breath noise in a woodwind instrument, the percussive noise of a piano hammer, and so on. Many researches have tried to decompose timbre into different attributes or descriptors. Robert Erickson ([16]) defined the five parameters initially described in 1968 by J. F. Schouten ([44]), as:

- The range between tonal and noiselike character
- The spectral envelope
- The time envelope in terms of rise, duration, and decay (ADSR—attack, decay, sustain, release)
- The changes both of spectral envelope (formant-glide) and fundamental frequency (micro-intonation)
- The prefix, or onset of a sound, quite dissimilar to the ensuing lasting vibration

Also, Erickson provided a table of subjective experiences and related physical phenomena based on Schouten's five attributes (see Table 3.1).

Moore ([38]) states that the distribution of energy over frequency is one of the major determinants of the quality of sound or its timbre (even though fluctuations over time can play an important role as well, as well as the relative phase of the

components), allowing to look at the overall distribution of spectral energy of the sound in order to have a first, rough idea regarding its timbre. Unlike pitch or loudness, timbre is a multidimensional variable that can not be measured on a single scale. Plomp ([41]) found that timbre was related to the relative level produced by sound in each critical band. Therefore, it could be possible that the dimensions of the variable timbre reach the number of critical bands, which is 37 ([38]).

Timbre judgements are highly subjective and connected to many different variables: the nature of the sound, its spectral content, the auditory sensation resulting from the perception of the sound, and, not less importantly, the gap between a sensation and the words used to describe it. Descriptive terms, in fact, can cause problems and ambiguities because they are, by their nature, describing subjective sensations. Usually, in order to describe timbres in psychoacoustic tests, listeners are asked to rate numerically sounds on scale of which the extremes are opposite terms, such as "bright" and "dark". Results are then averaged in order to remove, as much as possible, the effect of the subjectiveness of terms.

Timbre perception can also regard environments and rooms, because, as already explained, the reproduction of a sound in a room is altered both in the frequency and temporal domain, causing a change in timbre of the reproduced sound. Two important mechanisms connected to reflections are able to change the perceived timbre in a room ([50]):

- Acoustical interference (constructive and destructive) caused by the difference in phase of the direct and reflected sounds at the ears. Wether this is annoying or audible depends mostly on the number of reflections.
- Repetition of sound events, created by reflections. To a certain extent, this is perceived to be pleasant, as a room that is too damped can make the listener feel uncomfortable.

In [36], Wankling and Fazenda developed, through a number of psychoacoustic tests, a vocabulary of terms that could describe the timbre characteristics of sounds. Ferroni ([37]) also worked with these terms in his thesis work. This research uses their results in order to ask listeners for differences between sounds with respect to timbre parameters, also enriching the vocabulary through specific questions in the second psychoacoustic test.

3.4 Audibility of Modes and Acoustic Interference

As already introduced, resonance modes and acoustic interference are a very important factor that shapes the frequency response of a room at low frequencies, modifying the sounds that are played in the environment. It is important to briefly recap the concepts that relate this phenomenon to the way and extent to which humans are able to perceive it.

3.4.1 Comb Filtering

One of the causes of changes in timbre perceived in rooms is comb filtering, which is the result of phase offsets in the summation of a sound with another version of itself

that is slightly delayed, that cause constructive and destructive interference. This is the case when the direct sound emitted by the speaker is summed, at the listener's ears, with the early reflection of the same sound, after bouncing on the walls, which produces a slightly delayed, lower in amplitude and partially attenuated version of the original sound. The combination of the direct sound with the slightly delayed version of itself causes a destructive interference at certain frequencies, making the result look like a comb. The amount and position of the cancellation depends, of course, on the delay quantity (related to the room's dimensions), amount of damping, frequency content, and other variables.

Intuitively, destructive interferences occur when the direct sound and the reflected sound are out of phase (one half wavelength of difference). Therefore, the first frequency that will feature destructive interference will be the one at which the period is twice the delay ([50]):

$$Dipfrequency_1 = \frac{1}{period} = \frac{1}{2 * delay(seconds)} \tag{3.2}$$

The higher order cancellation frequencies will be the ones with an odd number of half wavelengths in the delay interval:

$$Dipfrequency_N = \frac{N(oddintegers)}{(2 * delay)} = \frac{1, 3, 5, 7, etc.}{(2 * delay)} \tag{3.3}$$

Conversely, the frequencies at which the peaks will occur will have a wavelength that is always multiple of the delay:

$$Peakfrequency_N = \frac{N(allintegers)}{delay(seconds)} \tag{3.4}$$

If the delay between the direct and reflected sound is longer (bigger room), the whole spectral pattern is shifted down in frequency. An example of the shape of a comb filter is shown in Fig. 3.3 on a logarithmic scale. With this scale, the appearance of the filter does not show that all peaks have the same size, however it is more representative of the way that we hear sound; above around 200 Hz, peaks become so near one to another that their contribution can be almost non-significant. However, at lower frequencies, it is indeed very important.

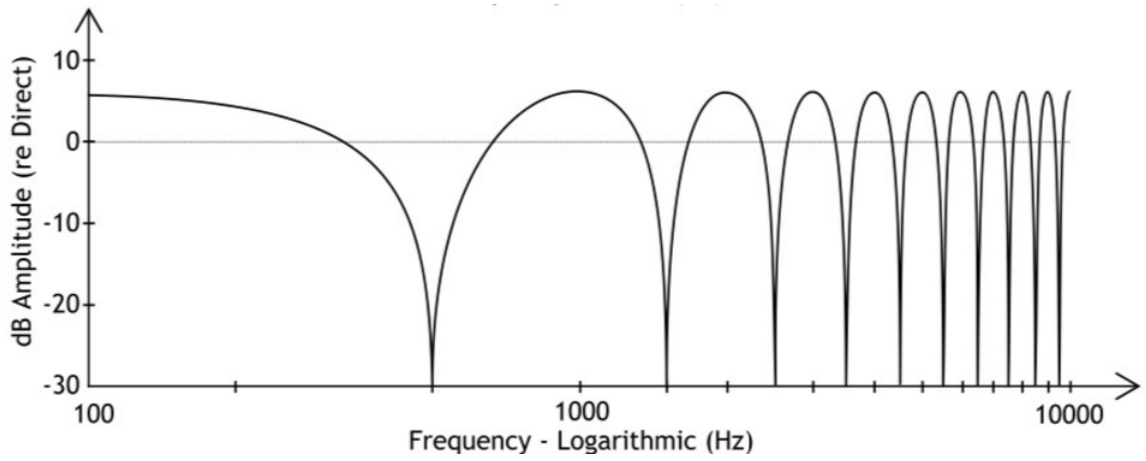


Figure 3.3: Example of the typical shape of a comb filter

It is important to remind the reader that this type of cancellation happens when the sounds have reached their steady state level. For transient sounds (and for the transient part of sustained sounds) the behavior is less predictable, and again depends on the amount of delay between the two signals ([50]). This is very important for this research, because, as will be made clear in chapter 5, in those frequency areas where there are cancellations, the steady state level is heavily affected, while both initial and final transients have different behavior.

3.4.2 Audibility of Resonances and detection of Q Factor changes

In sound reproduction, resonances should be avoided, as they alter the timbre of the sound that is played in the room, adding coloration. Playback systems should, in fact, accurately reproduce the content of the audio file rather than adding their own contribution. A frequency response peak can be evidence of a resonance, or resulting from acoustical interference between two transducers (such as a crossover region).

The simplest spectral deviation from the ideal, flat spectrum, is a "tilt". This is not a problem, since it is subject to adaptation: the ear usually gets used to it ([50]). For more serious problems such as peaks and dips, lots of experiments have been conducted during the years. As an example, Buchlein ([10]) performed the first tests regarding this problem, using headphones, concluding that dips are less noticeable than peaks, and that narrow interference dips are the least noticeable of all. On the other hand, wide peaks and dips are easier to detect than narrow ones. He also noted how they were difficult to detect with particular test sounds such as solo instruments, while they were audible when the test tone had a significant spectral content on the problematic frequencies.

If a dip were to have the same shape of a peak, only inverted, this would point to the presence of something that functions as a powerful resonant absorber of energy, which is a very rare occasion. More often, the cause of a dip is a destructive acoustical interference, in which case the dip will have the appearance of a very sharp, possibly very deep, dip at the frequency where the interfering sounds cancel ([38]).

More experiments by Toole and Olive ([21]) concluded that continous sounds are more revealing of resonances than isolated transient sounds and that humans appear not to be very sensitive to the ringing in the time domain at frequencies above around 200 Hz, but more to the spectral feature. However, at low frequencies and depending on the test tone spectral content, humans can hear both the spectral behavior and/or the ringing. Further analysis show that we are able to perceive the presence of very narrow (high-Q), low-amplitude, spectral aberrations.

If a tone burst at a fixed frequency drives a system with different Q-values for the same frequency, an interesting behavior can be seen by observing the temporal evolution of the response at the threshold of noticeability of the resonance ([50]). The test tone needs some time to drive the high Q resonance to its steady state value. This makes these resonances higher in amplitude than lower-Q resonances; for them to be energized, a musical tone must be at the correct frequency for a sufficient time, and that is a probabilistic event ([38]). After the sound is stopped, a longer ringing appears. The system with a lower Q resonance responds more quickly

to the tone burst, reaching full amplitude in a few cycles and decaying faster. One reason why we can detect medium-Q resonances having lower amplitudes than high-Q resonances is that it takes only a few cycles of signal to get them to respond to full amplitude. Since they have larger bandwidth, the frequency can be also not exactly centered on the resonances, even more so with low Q resonances.

Neither the measured amplitude of the spectral problem nor the duration of the ringing is a direct correlate of what we hear. In fact, the ability to hear a resonance strongly depends also on the ability of the driving signal to excite it (this is the reason why lots of attention will be put into designing and developing the psychoacoustic test for this research). The duration of the ringing is significative only at low-bass frequencies. Therefore, waterfall diagrams should be taken with a bit of caution: they can evidence the presence of resonances, but they don't correlate directly to their audibility.

An important fact should be highlighted: the amplitudes of the resonances shown in frequency responses are the steady-state measured changes in the playback system caused by the presence of the resonances "that have been adjusted to the detection-threshold level while listening to different kinds of program" ([50]). This value is not the amplitude of the output when listening to musical program material, since music is definitely not a steady state signal. Therefore, the amplitude of the output is likely to be lower. This is a key concept of this research work, that inspects the cases in which the frequency response fails to describe the auditory sensation at lower frequencies.

An early work by Olive ([7]) suggested that modal detection in the steady state is inversely proportional to modal Q, while it is directly proportional when using square pulses. Further tests by Fazenda ([14]) seem to generate results that agree with those obtained by Olive using square pulses, while contradicting those obtained by using continous noise signals.

Karjalainen ([51]) found that, while it is of primary importance to perform magnitude equalization, long modal decay times should be treated as well; however, at very low frequencies, the detection threshold for decay times increases, and he suggested that at very low frequencies long decays could not be perceived if the magnitude response is equalized well enough.

Fazenda ([14]) inspected the threshold value of the Q-factor for detectability. Results indicates that reductions to the Q factor below a value of about 16 are subjectively imperceptible, suggesting that any additional room treatment that reduces this value may be redundant. The detection thresholds obtained are 16 ± 5.4, 10 ± 4.1, and 6 ± 2.8 for comparison with sounds of reference Q values of 1, 10, and 30, respectively. Translating these values, it has been found that decay rates of 1 second or less at 32 Hz would be unnoticed by listeners. This decay detection threshold decreases rapidly with increasing frequency

Furthermore, [7] suggested that the reverberation at higher frequencies could impact the perception and detection of lower frequency resonance modes. One of the psychoacoustic tests of [37] seems to deny this dependance, even though more tests should be performed to be sure.

Finally, some of the results obtained by Ferroni in his research work ([37]), which was performed with the same impulse responses as those used in this thesis, include (with regards to the perception of modes):

- If resonances are strong enough, they affect the perception of onsets for sounds

repeated with sufficient speed, causing attacks to be less defined, damped or distorted

- Resonances seem to be more perceivable for sounds played by pitched instruments, with notes of sufficient duration
- Resonance modes always introduce a degradation of listening quality, producing a not faithful sound playback and unbalancing frequency response of the environment
- The presence of instruments also at higher frequencies doesn't mask problems due to resonance modes at low frequencies

The complete list of Ferroni's results can be found in chapter 1.3. It is interesting to note that the presence of room modes introduces a change in timbre and perceived listening quality if compared to the "dry" sound (the one that is not convolved with the room impulse response, or played on headphones). This could mean that different resonances may introduce different levels of degradation for sounds played at the resonant and nearby frequencies. This hypothesis is inspected in this research work.

Chapter 4

Room Analysis Techniques

4.1 Impulse Response

The impulse response is the signal that contains all the information about how the room reacts to sound. Starting from the IR, many important information can be derived regarding sound propagation in the environment. As already explained, the impulse response is also a function of the position of both source and receiver, and varies greatly with the position of both inside the room. Every room can be considered a LTI (linear time invariant) system, in which the superposition effect holds true. Therefore, the scenario is the one depicted in fig. 4.1, where x(t) is the input in the temporal domain, y(t) is the output, and h(t) is the impulse response. X(f) is the spectrum of the input signal, Y(f) the spectrum of the output signal and H(f) the frequency response of the system.

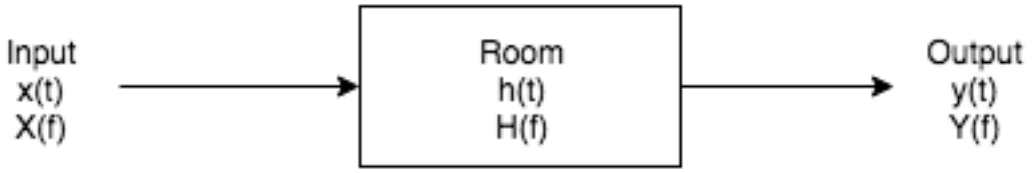


Figure 4.1: Room as a Linear Time Invariant system

An addictive noise component is also implicit, but is not reported for simplicity in this brief recap. In this scenario, two corresponding equations can be written, that describe the input-output relations

$$Y(f) = X(f) * H(f) \tag{4.1}$$

The previous equation ([24]) relates the spectrum of the output signal to the spectrum of the input signal by means of the frequency response; the following equation instead, relates the temporal behavior of the output signal to the one of the input signal, by means of the system's impulse response.

$$y(t) = x(t) \bigotimes h(t) \quad = \quad \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau \quad = \quad \int_{-\infty}^{\infty} x(t)h(t-\tau)d(-\tau) \tag{4.2}$$

The symbol $\bigotimes$ indicates the linear convolution operation. In eq. 4.2, the third and fourth terms explain how linear convolution is performed.

Once the impulse response of a room has been measured, the process of playing a sound in the room can be simulated by convolving the sound with the room's impulse response. The output of this operation is very similar to the measured response in the real room, at the same position of the receiver used to capture the impulse response and with the same source conditions.

Furthermore, as previously introduced, H(f) is actually the fourier transform of the impulse response, highlighting the strict correspondance between the time and frequency domains. From 4.2, it is clear that the impulse response is also the output of the system when the input is the so-called dirac function δ (a generalized function which is equal to 1 in the 0 position on the time axis and null everywhere else, so that its integral from minus infinity to plus infinity equals one):

$$y(t) = \delta \bigotimes h(t) \quad = \quad \int_{-\infty}^{\infty} \delta(t-\tau)h(\tau)d\tau \tag{4.3}$$

So, if the system is excited with a Dirac δ, or an approximation (impulsive signal with a very broad spectrum) the output is the impulse response, or an approximation of it. This can sometimes be quite difficult, so another family of methods for measuring the IR exists. They are based on the possibility of inverting the input function in the time domain, obtaining the inverse filter (a function that, convolved with the original signal, results in the δ function:

$$x^{-1} \bigotimes x = \delta \tag{4.4}$$

Therefore, it is possible to obtain the impulse response by applying the inverse filter to the output:

$$y \bigotimes x^{-1} = x \bigotimes h \bigotimes x^{-1} = h \bigotimes x \bigotimes x^{-1} = h \bigotimes \delta = h \tag{4.5}$$

The Signal-To-Noise ratio (S/N) is improved by taking multiple synchronous averages of the signal recorded in the environment, usually directly in time domain, before attempting the deconvolution of the system's impulse response. In order to apply these techniques, the spectrum of the input signal must be as flat as possible, otherwise its non-flatness would lead to a distorted impulse response [23].

4.2 Impulse Response Measurement Techniques

This section briefly describes some of the techniques used to measure an impulse response. This section only aims at giving an overview of the most common methods; if the reader is interested to learn more, some suggested reads are: [18], [19] [28], [27], [11]. For a comparison of these and more methods and their pros and cons, see [48]. It is obvious that every method has its tuning parameter and some preparation for each method is required in order to tune the system and deliver the best possible results. Besides the methods that use pulsive sources, the more complex methods (MLS, TDS, ESS) all follow the measurement scheme depicted in fig.4.2 , while changing the nature of the test signal and the type of post processing.

It is obvious that every method has its tuning parameter and some preparation for each method is required in order to tune the system and deliver the best possible results.

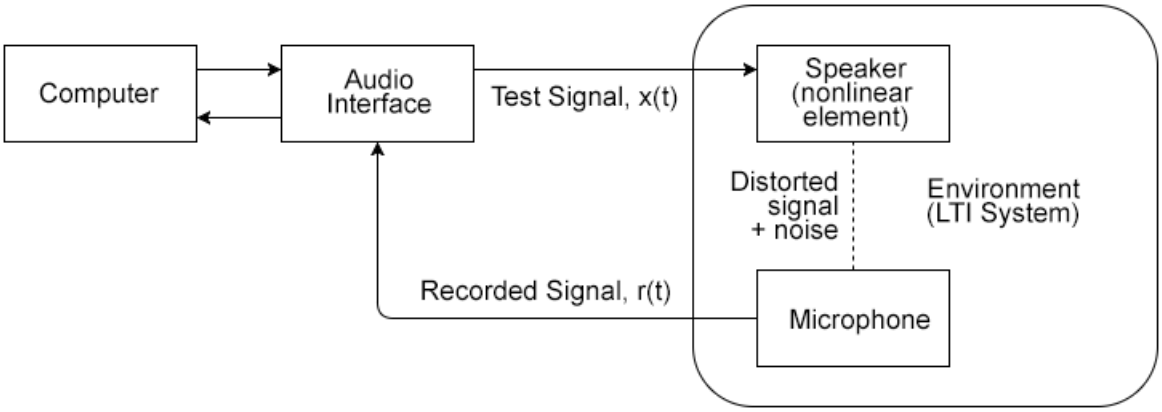


Figure 4.2: Scheme for IR measurements with test tones

4.2.1 Pulsive Sources

Using pulsive sources is the most basic method for measuring an impulse response. By exciting the system with a pulsive signal as close as possible to δ (very short pulsive sound, with a very wide spectrum range and flat spectral content), the actual recording of the room's response to this sound can be used as the impulse response without needing further post processing. This method is really simple but it carries many possible problems: first of all, it is not easy to realize such a signal, because the sound has to be really short and louder than the noise floor by tens of deciBels. Some of the approximations to the δ functions include the explosions of big balloons and gun shots. These sounds are not always completely reproducible (for example, because of slightly different dimensions of the air balloons). This makes it problematic to take different measurements of the same environment. Another downside to this method is that the spectrum of the source sound is not completely flat, exciting less frequencies than the human hearing range. This is not really problematic if the impulse response is used to derive some acoustic parameters, but the heavily uneven spectrum makes these measurements unusable as numerical filters. Despite these downsides, this method is still widely used thanks to the easiness of realization that makes it possible to have a rough idea of the acoustical behavior of the space [23].

4.2.2 MLS Method

This method uses, as a test tone, the MLS (Maximum Length Sequence) signal, a pseudo random binary signal that uses a shift register with a period of $L = 2^{N-1}$, where N is the number of slots in the shift register ([18]). This signal is similar to a square wave, but the length of the "1" and "0" regions is variable. The spectrum of this signal is flat except for its DC component. Also, its autocorrelation function yields an impulse signal and the cross-correlation function of a system's response to an MLS with the MLS itself is the system's impulse response ([23]). This signal is emitted through a loudpseaker and let reverberate in the room, while a microphone captures the resulting sound. The recording has to be processed in order to extract the system's impulse response, and thanks to the properties of the MLS signal, the Fast Hadamard Transform can be used (described in [18]). Because of the use of circular operations, the MLS technique deliverse the periodic impulse response $h'(n)$,

that is related to the impulse response by the formula:

$$h'(n) = \sum_{l=-\infty}^{\infty} h(n + lL) \tag{4.6}$$

where L is the length of one period. This formula shows the problem with the MLS method: if L is shorter than the length of the impulse response to be measured, it is possible that time aliasing errors generate ([23]). Each MLS sequencehas a strongly erratic phase spectrum, with a uniform density of probability in the $-\pi$, π range. Therefore, the phase spectrum of any component of the output signal which is not correlated with the MLS input sequence (disturbing signals) is randomized, leading to a uniform repartition of the disturbing effects along the deconvolved impulse response instead of localized noise contributions along the time axis ([48]). One of the problems with the MLS method is that distortion artifacts, more or less uniformly distributed along the deconvolved impulse response, might be present. This phenomenon is caused by the non linearities in the measurement system, especially the loudspeaker. This problem can be reduced by optimizing the measurement parameters, or using a variant of the MLS method, called IRS (inverted repeated sequence) method. The IRS signal is closely related to the MLS one and the deconvolution operation is the same ([48]).

4.2.3 TDS Method - Linear Sine Sweep

The TDS (Time Delay Spectrometry) method is also known in Japan as "stretched pulse" and in Europe as "chirp". It was invented in 1967 by Richard Heyser ([28]), and the idea was based on using a tracking filter directly in the frequency domain. His initial goal was to measure the response of an electro-acoustical system (such as a loudspeaker) by measuring just the initial part of the impulse response, related to the direct sound. To do so, a bandpass filter tuned to a given frequency could be used, leaving the system on only for the time it takes for the wavefront to reach the microphone. This would have been repeated for all frequencies, re-tuning the bandpass filter. It is not feasible to do this frequency by frequency, but a special signal that spans all frequencies fastly can be used instead:

$$x(t) = \sin\left[\phi_0 + 2\pi\left(f_0 t + \frac{k}{2}t^2\right)\right] \tag{4.7}$$

Eq. 4.7 is the equation of the sinusoidal linear chirp. Fig. 4.3 shows this signal in the time domain.

In a linear chirp, the instantaneous frequency f(t) varies linearly with time:

$$f(t) = f_0 + kt \tag{4.8}$$

where f_0 is the starting frequency (at time t = 0), and k is the rate of frequency increase or chirp rate, defined as:

$$k = \frac{f_1 - f_0}{T} \tag{4.9}$$

The energy in a varying frequency signal at any frequency is proportional to the time spent at that particular frequency; therefore, the energy of a linear sine sweep

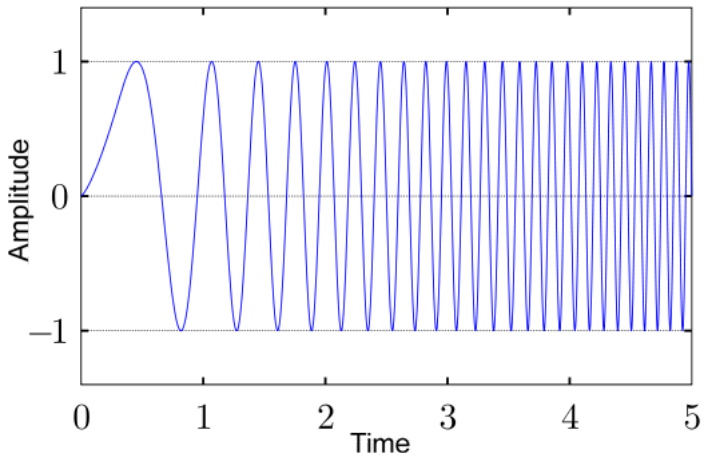


Figure 4.3: Linear Sine Sweep (Chirp) signal in the time domain

as a function of time is given by:

$$E(t) \propto \frac{T}{f_1 - f_0} \tag{4.10}$$

The energy is therefore constant for all frequencies. The inverse filter is the linear sine sweep signal reversed along the time axis; Usually, f_0 and f_1 are chosen to be respectively 20 and 20.000 Hz in order to span the whole human hearing range, and the duration of the signal is usually set to 15 seconds ([27]).

The impulse response can be obtained by simply convolving the measured microphone signal, $r(t)$ with the inverse filter, $s(-t)$:

$$h(t) = r(t) \bigotimes s(-t) \tag{4.11}$$

A linear sweep has a "white" spectrum, and consequently provides a signal-to-noise ration which is not very good at very low frequencies. In order to overcome this problem, Hidaka and Beranek used a different version of this technique called "double stretched pulse" using two different sweep rates: a slower one at low frequencies and an higher one for medium and high frequencies ([18]).

4.2.4 ESS Method - Exponential Sine Sweep

This technique is, as a matter of fact, a variant of the TDS method, with a different test signal. Both MLS and TDS methods are based upon the assumption of a perfect LTI system, and generate problems when these assumptions are not true; Angelo Farina et. al ([18], [20], [19], [17]) have studied, developing the work of other authors like Gerzon and Griesinger, how the use of exponential sine sweep signals instead of linear sweeps makes it possible to deconvolve simultaneously the linear impulse response of the system, and separate impulse responses for each harmonic distortion. As pointed out by Farina, the signal played by the loudspeaker is composed of harmonic distortions. It is quite difficult to separate the linear part (reverberation in the Impulse response) from the nonlinear part (distortions) . Therefore the response of the system can be considered as the composition of additive gaussian white noise,

and a set of impulse responses, each of them being convolved by a different power of the input signal ([48]).

The exponential sine sweep method produces advantages respect to other techniques: in particular, it produces better S/N ratio and the presence of peaks that contaminate the late part of the MLS responses is absent, allowing for more precise measurements of the system's linear impulse response even if the loudspeaker is working in a non-linear region ([17].

The corresponding time-domain function for a sinusoidal exponential chirp is given in ([18]):

$$x(t) = \sin\left[\frac{\omega_1 T}{ln(\frac{\omega_2}{\omega_1})} * \left(e^{\frac{t}{T} ln(\frac{\omega_2}{\omega_1})} - 1\right)\right] \tag{4.12}$$

This is a sweep which starts at the initial radian frequency ω_1, ends at the final radian frequency ω_2, over the course of T seconds, and where the istantaneous frequency varies exponentially as a function of time.

The signal, in the time domain, is shown in fig. 4.4. It important to focus on the fact that this signal does not have a flat spectrum, because the instantaneous frequency varies slowly at low frequencies and faster at higher frequencies, resulting in a pink spectrum (falling down by -3 dB per octave). The inverse filter has to compensate for this, therefore an amplitude modulation is applied to the reversed sweep signal so that it increases by 3 dB per octave ([17]).

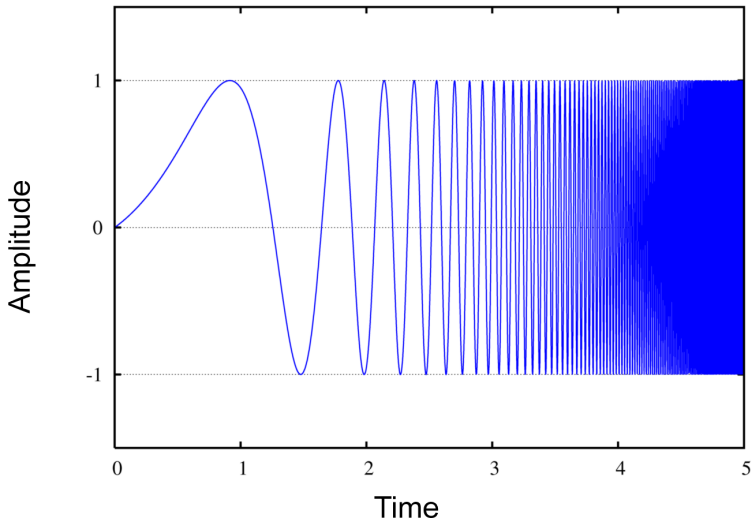


Figure 4.4: Exponential Sine Sweep (Chirp) signal in the time domain

This signal, having constant amplitude, is played through the loudspeaker in the environment; the room response is recorded through the microphone. It is often possible to see some distorted harmonic components, which are caused mainly by non-linearities in the electro-mechanical transducers. Using linear convolution allows time-aliasing problems to be avoided: by the use of linear deconvolution, it is possible to avoid that the anticipatory harmonic distortion responses (which occur earlier than the linear response) fold back inside the time window, contaminating the late part of the impulse response. This way, the linear impulse response is assured

exempt from any non linearity, and measurement of the harmonic distortion at various orders can be performed.

The Inverse Filter f(t) is generated in the following manner ([48]):

- The Logarithmic Sweep (which is a causal and stable signal) is temporally reversed and then delayed in order to obtain a causal signal. This time reversal causes a sign inversion in the phase spectrum. By convolving this result with the initial SineSweep, a signal with perfectly linear phase is obtained (corresponding to a pure delay), but it will introduce a squaring of the magnitude spectrum.

- The magnitude spectrum of the resulting signal is then divided by the square of the magnitude spectrum of the initial SineSweep signal.

After the inverse filter, f(t), is obtained, the impulse response can be computed with:

$$h(t) = r(t) \bigotimes f(t) \tag{4.13}$$

4.2.5 Comparison of techniques

In [48] a comparison of some of the most used techniques for measuring impulse responses has been carried out. Some of the most important results are reported in the following:

- The MLS (IRS) method seems the most interesting method when the measurements have to be performed in an occupied room or in exterior due to its strong immunity to all kinds of noise (white, impulsive or others), its weak optimal output sound level and its timbre (white noise is more supportable and easily masked out than sweep signals). However, its major drawback lies in the tedious calibration that has to be carried out to obtain optimal results and in the appearance of spurious peaks ("distortion peaks") due to the inherent non linearities of the measurement system.

- The Time-Strectched Pulses method avoids the appearance of the distortion peaks. However, the remaining non linear artifacts can possibly be superimposed with the deconvolved "linear" impulse response. The presence of a residue of the excitation signal in the deconvolved impulse response is a result of such superposition problem. This residue can be almost completely eliminated if a precise calibration of the measurement parameters (mainly the output level) is realized. However, its timbre and the high value of the optimal output signal level needed to mask out the ambient noise makes it unusable in occupied rooms.

- The perfect and complete rejection of the harmonic distortions prior to the "linear" impulse response, their individual measurement and the excellent signal-to-noise ratio of the SineSweep method make it the best impulse response measurement technique in an inoccupied and quiet room. Moreover, unlike the preceding methods, it does not necessitate a tedious calibration in order to obtain very good results (no compromise between the signal-to-noise ratio and the superposition of non linear artifacts in the room impulse response). However, as for the Time-Stretched Pulses method, the SineSweep technique is not recommended for measurements in occupied rooms.

4.3 Fourier Analysis and Modal Analysis

4.3.1 Fourier Analysis

The Fourier Analysis is the study of the way general functions may be represented or approximated by sums of simpler trigonometric functions. The process of decomposing a function into oscillatory components is often called Fourier transform, while the operation of rebuilding the function from these pieces is known as Fourier synthesis. As an example, in order to compute which frequencies compose a musical note, a Fourier transform of the sampled note can be performed; the same sound can also be created by summing up the different frequency components revealed by the Fourier analysis. Different types of Fourier Transform exists, and they deal with different cases; each specific transform is used on a set of signals, differentiating between continous and discrete, periodic and aperiodic ones. If the reader is interested on learning more on this subject, a suggested read is [24].

The process known as Fourier transform allows one to switch from time domain to the frequency domain, in order to visualize information (relative to both magnitude and phase) about the spectral content of the signal. The inverse transform allows to switch back from the frequency domain to the time domain, with a perfect correspondance and without losing any information. Fourier theory states that it is possible to describe a time function using a series of sine and cosine functions ([37]):

$$S_{f,n} = \frac{a_0}{2} + \sum_{-\infty}^{\infty} (a_k cos(kx) + b_k sin(kx)) \, \forall x \in \Re \tag{4.14}$$

If the function has a period of 2π and is piecewise continuos, it converges with Fourier's series.

A few properties of this transform include:

- The transforms are linear operators
- The transforms are usually invertible.
- The exponential functions are eigenfunctions of differentiation, which means that this representation transforms linear differential equations with constant coefficients into ordinary algebraic ones. allowing the analysis of a linear time-invariant system to be performed at each frequency independently.
- By the convolution theorem, Fourier transforms turn the convolution operation in the time domain into multiplication in the frequency domain, providing an efficient way to compute convolution-based operations.

Sometimes, Fourier transforms can have high computational complexity. Fast Fourier Transform (or FFT) is a computationally efficient method for computing such a transform on digital signals, working on finite length blocks of sampled data called FFT frames. This transform has been used in this work in order to derive the room's frequency responses from the impulse responses, through Matlab™.

With regards to this research work, the importance of the Fast Fourier Analysis lies in the possibility of accessing the information regarding the room's frequency response starting from the impulse response. The problem of this approach is that the frequency response shows the steady-state behavior of the system; however, as

will be made clear in the following chapters, this information is psychoacoustically meaningful mostly for long, sustained sounds. In music especially, sounds have also important information in their onset and offset transients, and when notes are really short, the steady-state level is not always reached. The FFT of an impulse response will show problems due to resonance modes in the highly non-flat low-frequency region of the frequency response. For this region, as already explained, the room is considered non-sabinian and the concept of reverberation times loses meaning. A modal analysis, concentrating on each specific mode, has to be carried out in order to analyze properly each resonance mode, as different frequency have vastly different behaviors in their temporal envelope.

Time and frequency domain measurement resolution

Before moving on, an important note should be done in regards to the plots that show both time and frequency domain information, such as waterfall plot. As a matter of fact, a drawback of the FFT is the trade-off that must be made between frequency and time resolution; the more accurately we want to measure the frequency content of a signal, the more samples we have to analyze in each frame of the FFT. However, the larger the frame, the less is known about the temporal events that take place within that frame. In other words, more samples require a longer observation time; but the longer the time, the less the sound over that interval looks like a sine wave, or something periodic, therefore being less well represented by the FFT ([1]).

Therefore, the information these plots carry is compromised in both time domain and frequency domain axes, and the compromise can be manipulated to favor one or the other, but not both. In other words, one can have high resolution in the frequency domain and sacrifice resolution in the time domain, or the reverse. All of this is most relevant at low frequencies ([50]).

The obtainable frequency resolution is equal to the inverse of the time duration during which a signal is observed, i.e., $1/T$. This is because in order to obtain a certain deterministic frequency resolution, one must be able to observe at least a full cycle (of duration T) of a sine wave at the frequency of interest. Conversely, the otainable time resolution is equal to the inverse of the bandwidth used to observe the signal at baseband, i.e., $1/f$.

In the FFT, windows of samples are analyzed together, before the window is moved to a group of consequent samples. 1,024 samples (1k) is usually used as the frame size for an audio FFT ([1]). At a sample rate of 44.1 kHz, 1,024 samples corresponds to 0.022 second of sound. Therefore, all the events that take place within that 0.022 second will be lumped together and analyzed as one event. Because of the nature of the FFT, this content is actually treated as if it were an infinitely repeating periodic waveform. The amplitudes of the frequency components of all the sonic events in that time frame will be averaged, and these averages will end up in the frequency bins. This is known as time smearing. If a better frequency resolution is needed, a bigger frame size has to be used. But this means that even more samples will be lumped together, worsening the time resolution.

On the other hand, if good time resolution is more important, the frame size should be smaller, ideally one sample. Unfortunately, with only one sample to analyze, no useful frequency information would result from the FFT analysis. A more reasonable frame size and one that is considered small for audio, such as 256 samples (a 0.006-second chunk of time), corresponds to 128 analysis bands, for a bin

width of about 172 Hz. While a 0.006-second time resolution is reasonable, 172 Hz is a very bad frequency resolution.

There is not a solution to this problem; the compromise is to balance the resolution in both domains based on the type of sound under analysis ([1]).

A visual example of this tradeoff can be viewed in [50], page 246, where a waterfall plot is shown with three different choices of frequency and time tradeoffs.

It was worth mentioning this relation between time and frequency resolution; this is still valid in the computation of the FFT in this research work. However, thanks to the way the AQT algorithm (the one used to analyze the rooms in this work) is constructed, this research did not have to deal with this problem after the computation of the frequency responses. In fact, the algorithm convolves a sine burst for even frequencies in the range 20-300 Hz (therefore, the spacing is 2 Hz) with the impulse response, allowing to have access to the precise response of all these frequencies on the same time axis. Therefore, both the EFT Plot and the Waterfall Plot generated by the expanded AQT algorithm do not suffer from this problem. Having analyzed tone bursts with a difference of 2 Hz, however, the missing frequency are most likely averaged when the Waterfall Plot is plotted. The algorithm will be described in more detail in the next sections.

4.3.2 Modal Analysis

As already explained, in the modal region (below the Schroeder frequency) the frequency response behavior and performance is dominated by the presence of room modes, and the sound field can not be considered diffuse. Modes behave differently, and they are not absorbed strongly as sounds that bounces on all surfaces; the fact that the incidence of sound waves for such frequencies is non-random also changes the absorption. This results in longer decay times at those specific frequency, meaning that a single decay time is not useful to describe the whole behavior of the room; instead several decay times are present, and the longer ones are related to the most problematic resonance modes in the room, causing a degradation in the listening quality and the perceived precision and definition of sound (this research aims also at inspecting this behavior). Acoustic treatment at low-frequency is very impractical and problematic, because it can be expensive (the cost of absorbers arises with decreasing frequency) and the size of the acoustic treatment itself can be too big for the room in which it has to be put.

Karjalainen et al. ([51]) found that at very low frequencies, longer decay times are allowed as the human ear is less able to perceive them, concluding that down to about 100 Hz the temporal properties of the modal decay are somewhat critical, while below this quantity, and even more so below 50 Hz, even very long decays (up to 2 seconds and above) may not be noticeable if the magnitude response is well corrected. The results of this research work seem to show otherwise, since the decay time of very low frequencies is perceived to alter the precision of the sound; however, it is important to point out that the frequency responses of the rooms analyzed in this research were far from being well equalized, in fact they were chosen exactly for their problematic nature.

After this recap on Fourier and Modal analysis, it is clear that an algorithm that is able to describe both the steady state response and the specific temporal behavior of the envelope of each frequency in the environment is needed. The so-

called Acoustic Quality Test is the appropriate tool for this analysis.

4.4 Acoustic Quality Test

The Acoustic Quality Test, also called AQT, is a new measurement technique of the acoustical quality produced by a sound system. Its output is a graphical representation of the dynamic response of the system to tone bursts at various frequencies, visualizing both the steady state frequency response, the transient response of the system, called "overshoot response", and other more advanced parameters. Its results have also been proved to be quite close to the perception of sounds in small environments, making it the best candidate tool for such an analysis.

The need for such a tool arises from the well known low-frequency behavior which is dominated by room modes. Furthermore, in small compartments such as cars, the reflected sound often arrives to the listener's ears well before the direct sound has finished playing, mixing the two sounds and therefore eliminiating the possibility of separating the direct sound from the reverberant field, as would be the case with large listening rooms. This makes the classical parameters such as reverberation time, clarity index and definition, less useful in these cases. Since the material reproduced in small rooms is often speech or music, both of which are highly non-stationary, the transient behavior of the system should be taken in consideration as well.

The first version of AQT was initially developed by I. Adami and F. Liberatore ([30]). This method can be considered as an evolution of the Music Articulation Test Tone (MATT) developed my M. Noxon ([40]). The first implementation of the AQT algorithm was then improved by A. Farina et al.([8]), first by creating its virtual counterpart (Virtual AQT), and then by developing the AQT2 method, which is based on the same principles but offers more solid results thanks to a different internal algorithm. The three versions of the algorithm will be described, explaining why they are so important to this research work.

In this thesis, the AQT 2 algorithm has been further developed in order to accomodate the measurements required to analyze a small room (since the original AQT algorithms were more focused towards little listening environments such as cars). If not stated otherwise, in the following chapters, the term "AQT algorithm" will be used with reference to the AQT 2 version developed by A. Farina et. al ([8]).

4.4.1 Original AQT Algorithm

The original AQT algorithm features a special test signal, which will be reproduced in the space to be assessed. The signal is composed of a series of sinusoidal bursts, each with duration of 200 milliseconds, intervalled by 66 milliseconds. The frequency is slightly augmented with each burst, with 2 Hz increasements from 20 to 300 Hz, 4 Hz increasements from 300 to 1000 Hz, and exponential increasement from 1 to 2 kHz. A zoomed in version of the test signal, showing two sine bursts, can be viewed in Fig. 4.5 ([8]). The very rapid amplitude transitions at the beginning and end of each burst are prone to produce some frequency-domain artifacts. So, the sequence of burst does not excite exactly one single frequency at a time, but rather excites it with a small filterbank with an aperture of approximately 1/6 of octave.

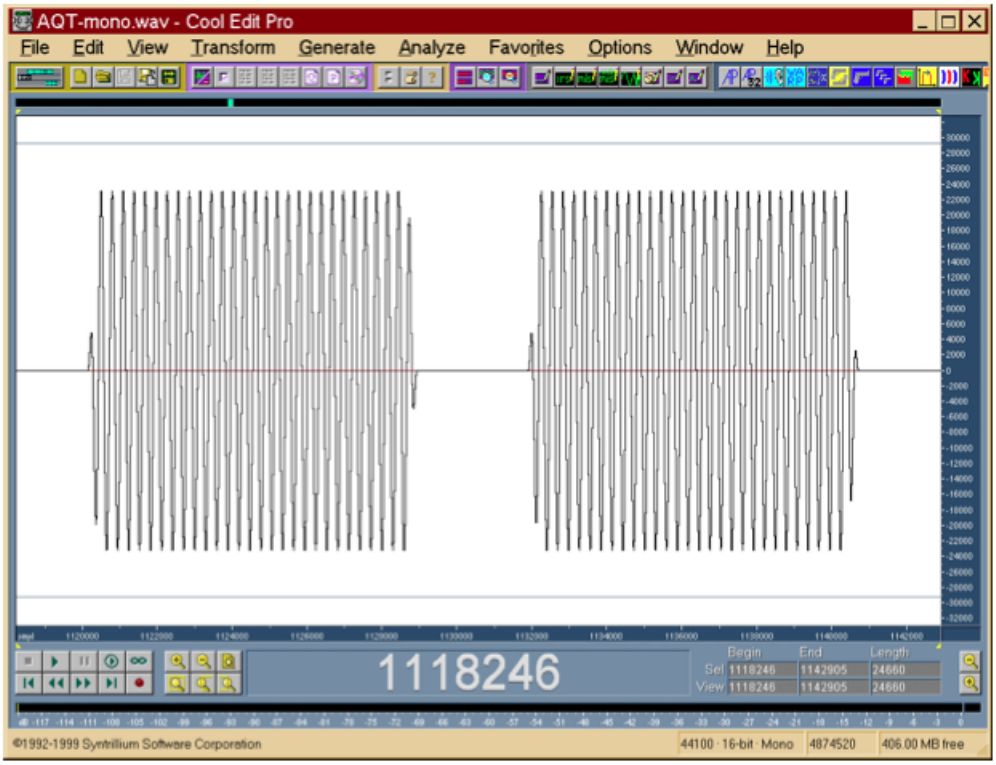


Figure 4.5: Two short sine bursts that are part of the original AQT test tone ([8])

The test signal is then played in the environment to be assessed. The instantaneous RMS level is measured at the listening position and a temporal recording is measured. The graphical plot of this level recording is the AQT measurement result: it is very simple and does not require further mathematical processing. However, a more meaningful version of this measurement can be found with some simple post processing, that allows the visualization of a level vs frequency plot, that can be seen in fig. 4.6 ([8]).

By zooming in on the frequency axis (in fig. 4.6), an interesting behavior can be seen, which is depicted in fig. 4.7 ([8]). At the left side of the figure, the sound which is directly emitted by the loudspeaker is substantialy in phase with the sound reflected by the room. In fact, the sound level first rises to a value of about 95 dB, then continues rising up until more than 97 dB, forming a response envelope which appears rounded. For slightly higher frequencies, the direct and reflected sounds have opposite phase: after the initial part of the burst, the RMS level is reduced when the reverberant sound field establishes, because of a destructive interference between reflections and direct sound. Two evident spikes (overshoots) can be seen at the beginning and end of the response envelope. These overshoots happen when only the direct sound is present (the beginning) and when only the reflected one is present (at the end) ([8]).

The numerical evaluation of AQT is based on the decay that happens in between the bursts; the human ear is subject to masking after hearing a loud sound, and after 66ms the masking curve is approximately at –20 dB ([8]). This means that 20 dB can be considered as an upper limit of subjectively perceivable amplitude modulation (also called "articulation"). If the articulation is lower than 20 dB, we are experiencing a "sustain" at the end of the note caused by reverberating energy

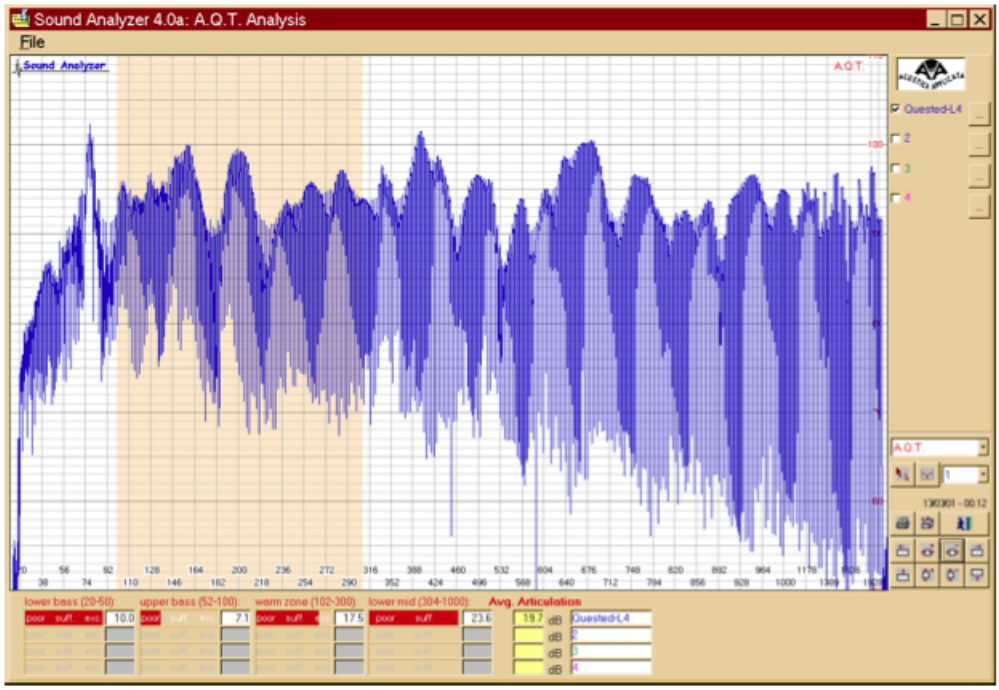


Figure 4.6: Results of the original AQT algorithm (level versus frequency) ([8])

attached after the end of the burst. Such short-term tail is very different from the longer tail tipically present in concert halls and big listening rooms, where the direct sound is followed by a gap of silence before the first reflections set in (tipically after 15-30 milliseconds), allowing the direct sound to be completed when the reflections arrive. This low term tail instead smears the transient and is detrimental to the dynamic perception of the sound. Therefore, modulation depth should be as large as possible, within the perceivable limit of 20 dB. The lowest level reached at the end of the burst depends, besides the decay of the burst, also on the attack of the subsequent one and the background noise.

Besides these, the original algorithm featured a user interface that could be used to inspect more details and features of the AQT results. In the original AQT analysis, the frequency range was divided into four main zones:

- Lower bass - from 20 to 50 Hz
- Upper bass - from 52 to 100 Hz
- Warm Zone - from 102 to 300 Hz
- Lower Mid - from 304 to 1000 Hz

In this research work, analyses will be focused on the 20-300 Hz area (lower bass, upper bass, warm zone) since it is the one where room modes are dominant and create the most interesting environment to test the hypothesis of this thesis work.

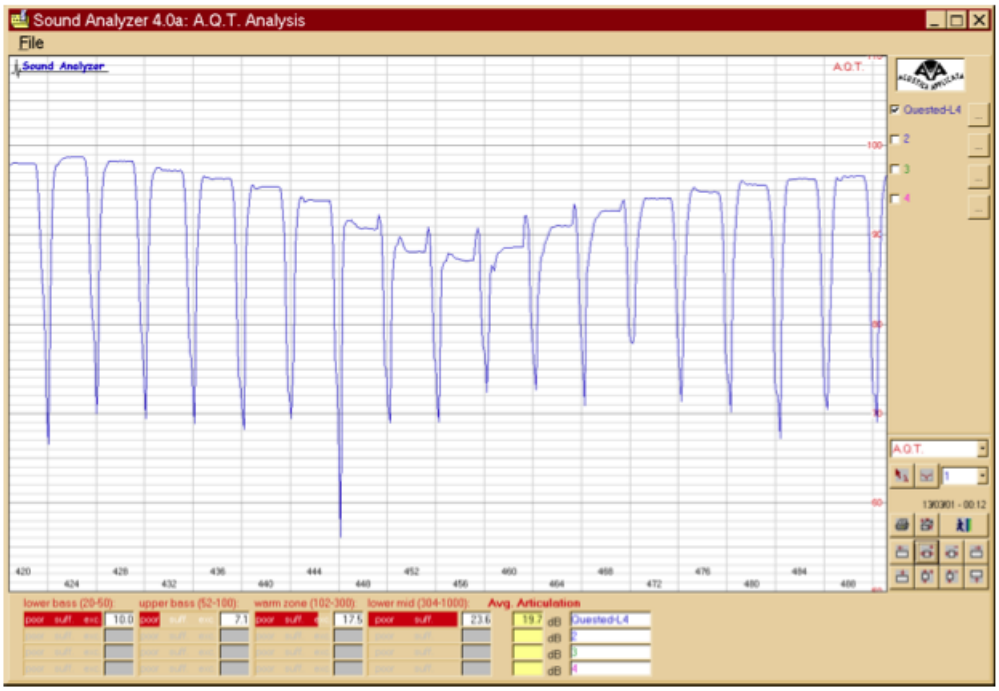


Figure 4.7: Zoomed version of fig. 4.6 ([8])

4.4.2 Virtual AQT

The first development of the original AQT method allowed to convolve the test signal with the previously measured impulse response of the environment under test, obtaining a realistic simulation of the system's response to the test signal. The original measurement procedure was quite long and sensitive to impulsive noise events; instead, with a measurement of the impulse response (with fast and solid techniques such as the MLS method or the sine sweep method), the AQT analysis can be performed on the synthetic wav file resulting from the convolution of the AQT test signal with the measured IR. Comparing the synthetic result obtained in the way just described to the measurement of the AQT result described in the previous paragraph, both by inspection and by listening to the sound, it seems that a difference can be found only in the silence section before and after the bursts, where the convolved file is missing background noise. The transients are carefully reconstructed, even in minor details.

Fig. 4.8 shows a comparison between these two signals. The profiles are close to each other, with some minor deviations which are, however, of the same order of magnitude of the variation that happens when repeating two "live" AQT measurements, so they can be considered non-significant. The convolution approach can be thought of as the natural complement of the AQT analysis, making the measurement faster and results more solid.

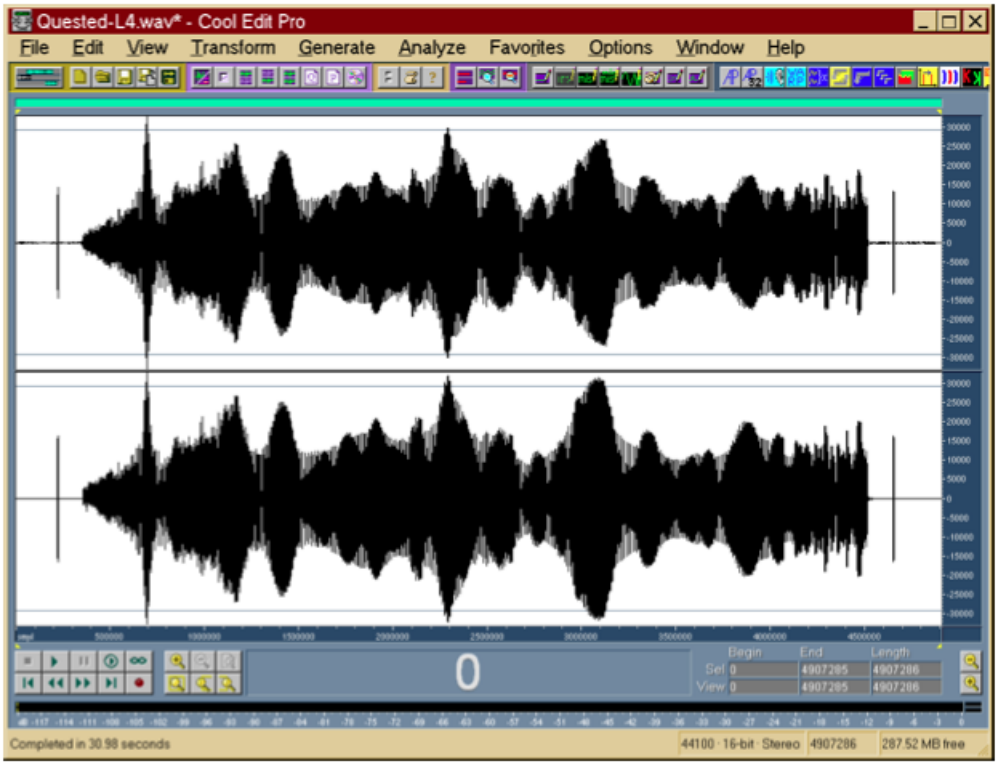


Figure 4.8: Comparison of the envelope of the sound obtained with the original AQT algorithm and with the Virtual AQT algorithm ([8])

4.4.3 AQT 2

AQT 2 is an even more advanced version of the AQT algorithm, created on a standalone application with Matlab™. Instead of creating an unique test tone, each tone burst is processed separately from the others, eliminating the problem of the interaction between the tail of one burst with the attack of the following one. Many separate bursts were generated (composed of 50 milliseconds of silence, 200 milliseconds of burst, and 300 milliseconds of silence) and then linearly convolved with the impulse response. The result can be seen in a sonograph (Energy Frequency Time plot) such as Fig. 4.9 ([8]), which shows the transient response of the system at all frequencies simultaneously with great detail. In the figure, a long reverberant tail can be seen at 74 Hz.

When clicking on the sonograph at a certain frequency, the response envelope is visualized (that it, the envelope of the temporal response of the system at that frequency). Certain frequencies, such as this one (pictured in Fig.4.10, [8]), show that the system was very slow in reaching the steady state, still increasing after 200 milliseconds, and generating a long decay after the sound's end. At other frequencies, such as the one depicted in Fig.4.11 ([8]), the system is very quick in following the burst shape, rising fastly to the steady state level and generating a short decay.

The same interesting behavior seen in the original AQT results appears in the regions where the phase the phase of the reverberant field is opposite to the phase of the direct sound: very evident spikes (overshoots) appear at the beginning and at the end of the burst. An example of this behavior can be seen in Fig. 4.12 ([8]).

The presence of these strong overshoot phenomena suggested the importance

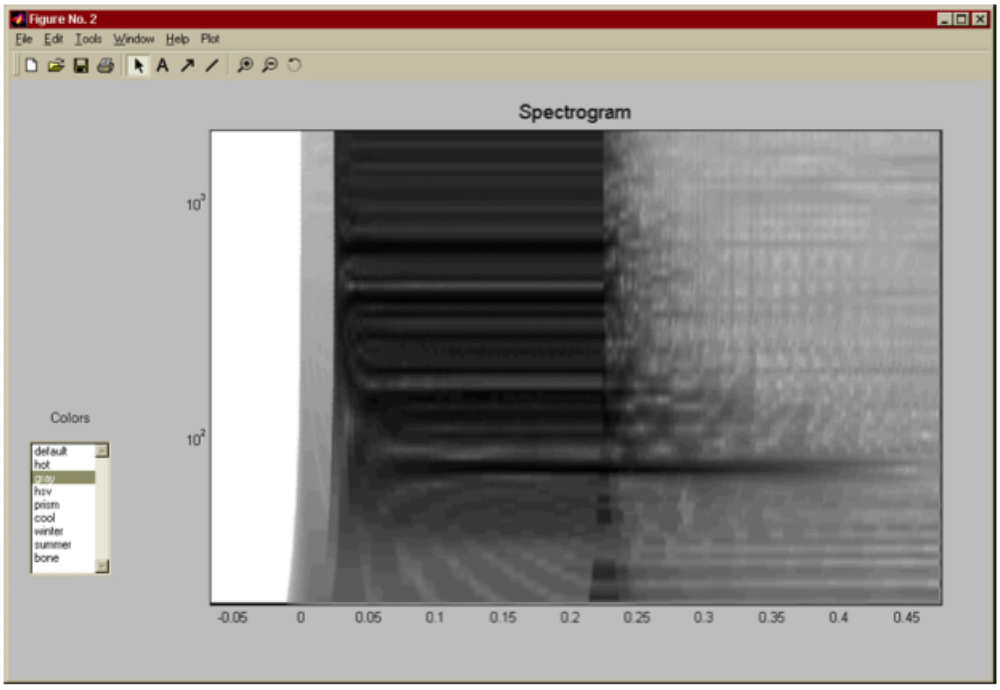


Figure 4.9: Sonograph of the transient response (EFT Plot) ([8])

of measuring and evaluating both the steady-state frequency response (which corresponds to the level in the central plateau between the two spikes), and the overshoot response, that is the maximum RMS level reached at any frequency in correspondance of the burst extremes. For each single burst four parameters were generated by a numerical analysis:

- Steady State level
- Maximum overshoot level
- Level after 33 ms
- Level after 66 ms

These values were plotted in the same graph (Fig. 4.13, [8]), showing all curves with respect to frequency. The 33 milliseconds curve was included as it was significative for really small environments such as cars. For our research, longer times should be considered as environments (small rooms) are bigger. On the room's resonance frequency, the decayed level after 66 milliseconds is almost coincident to the steady state level, and close to the overshoot level.

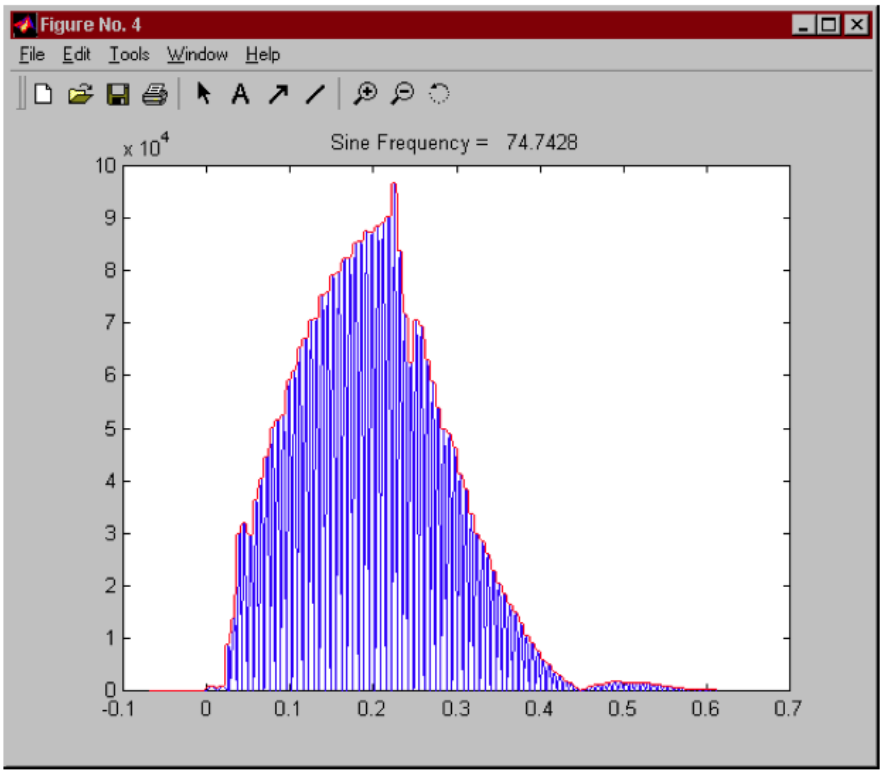


Figure 4.10: Response Envelope that does not reach its steady state, on a peak in the frequency response ([8])

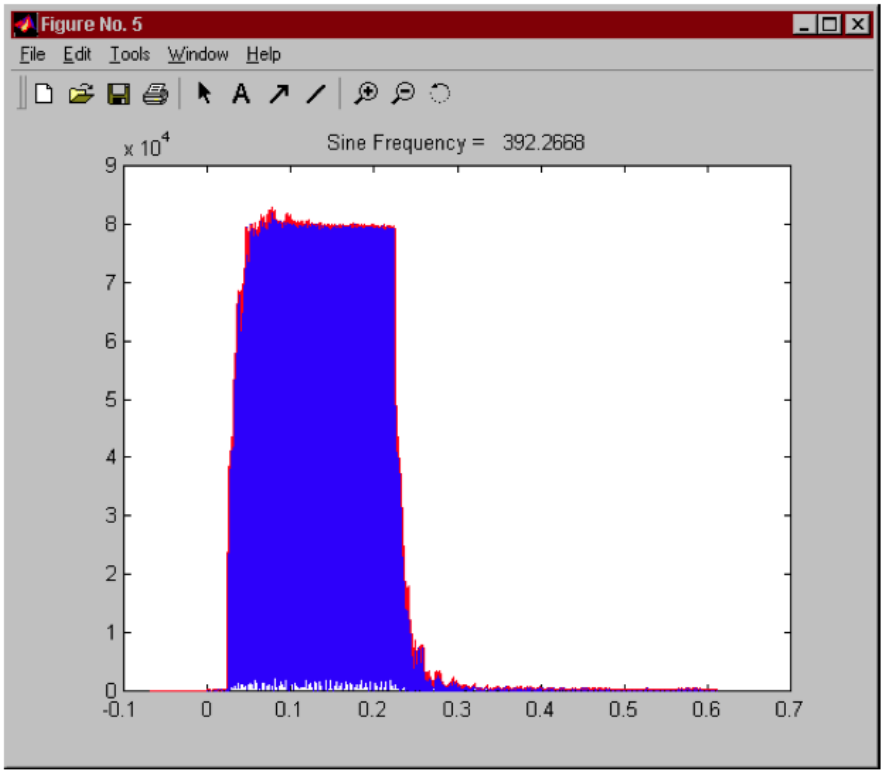


Figure 4.11: Response Envelope that reaches its steady state, on a peak in the frequency response ([8])

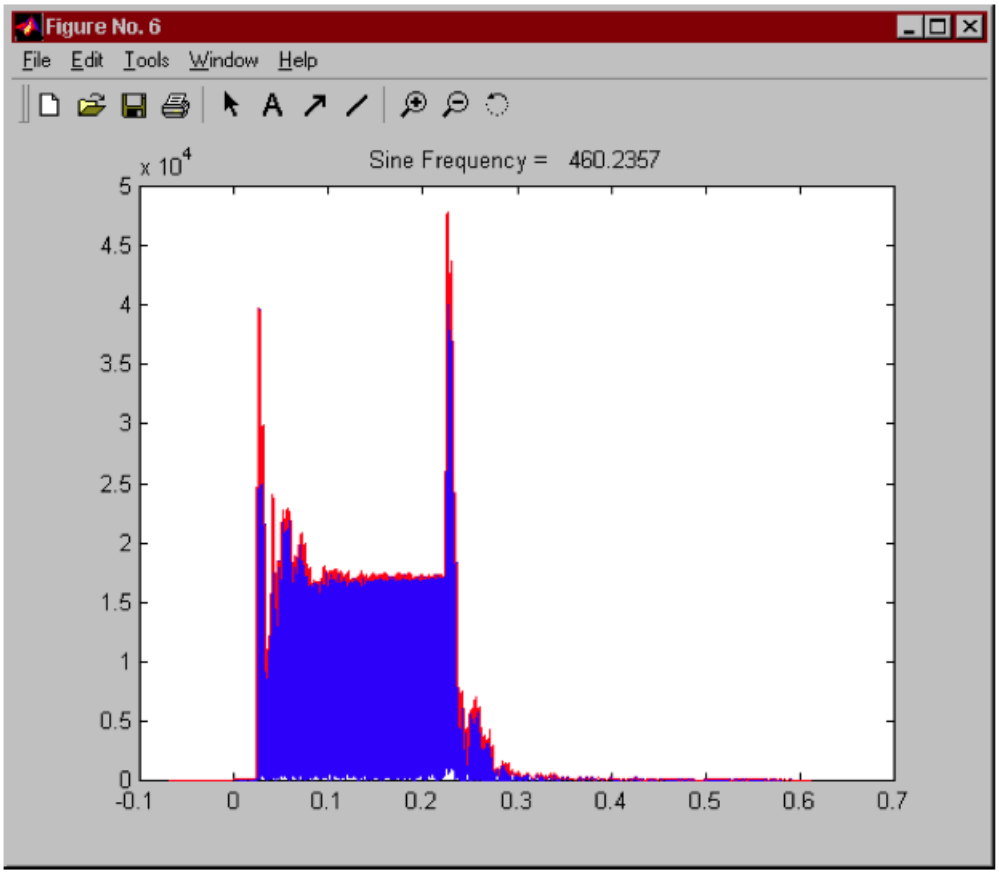


Figure 4.12: Response Envelope on a valley where there is destructive interference between direct and reflected field, showing overshoot behavior ([8])

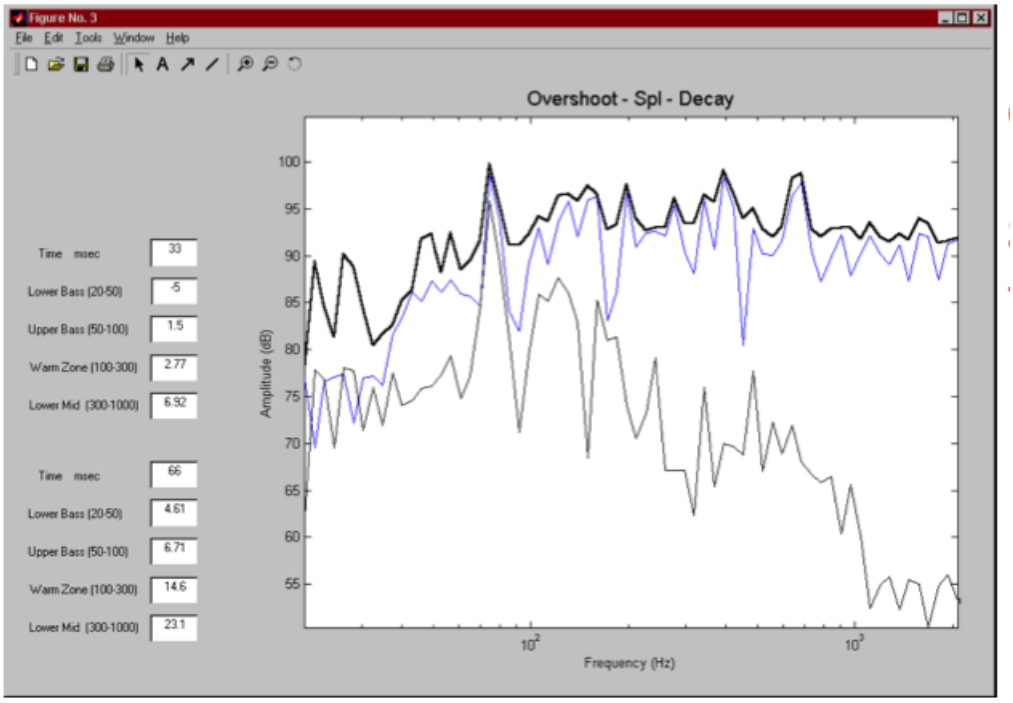


Figure 4.13: Steady State response, Overshoot response, Level after 33 and 66 milliseconds ([8])

Comparing the overshoot response and the steady state response with respect to frequency, the first appears to be much flatter than the second, and is close to the measured anechoic response of the loudspeaker. The subjective response corresponds with this flatness and the listeners perceive the sound of the loudspeaker as inherently flat and uncolored, even though the behavior in the listening room shows strong problems in the frequency response ([8]). Therefore, it seems that the ears are more sensible to the overshoot response than to the steady state response, suggesting that the overshoot curve can be better for describing the frequency response of a sound system composed by loudspeaker and small room. An informal listening test confirmed the results: testers were asked to choose which equalization they preferred, choosing between two instances of sounds, one with loudspeakers equalized with the inverse of the steady state response, and the other one with the equalization curve computed with the overshoot response. Testers liked better the second one, designed by flattening the overshoot response.

The following chapter presents and describes the further developments made to the AQT family of algorithms, and the analysis of eight impulse responses of real rooms using the new algorithm.

Chapter 5

Small Room Analysis

5.1 AQT Algorithm Development and Evolution

The AQT 2 algorithm, as presented in [8] was implemented in Matlab™ by the author. The script takes an impulse response as the input, as well as the burst length, tone test amplitude, initial frequency, frequency step and room dimensions as tuning parameters.

The algorithm creates the AQT test pure tone bursts, as defined in [8] and then convolves them with the Room Impulse Response, showing, for each frequency, the temporal envelope of their response. This response will be referred in the following sections as AQT Response Envelope.

Eight real room's Impulse Responses, professionally measured by "Suono e Vita" have been initially analyzed with three different duration of test bursts.

These analyses have been carried out from 20 to 300 hz with a step of 2 hz, as in the original AQT2 algorithm. The reader is invited to keep in mind that all behavior below the speaker's cutoff frequency (with reference to the speaker used to measure the Room Impulse Response) are subject to artifacts and may present unexpected behavior. Some developments were also added by the author, such as:

- Decay Time computation
- Further temporal behavior analysis
- "Room Slowness" and "Room Inertia" parameters
- Advanced Overshoot analysis

The following sections describe the algorithm, the developments and the results.

5.1.1 Block Schematics of the Matlab™ script

Fig. 5.1 shows the block scheme of the AQT algorithm as implemented by the author in Matlab™ with the aforementioned developments. On the right of each block, a brief overview of the inner processes describes what is performed in that stage.

The blocks "Room Slowness Computation", "Room Inertia Computation" and "Overshoot Analysis" were added after the first ("blind") psychoacoustic test and before the second ("focused") psychoacoustic test.

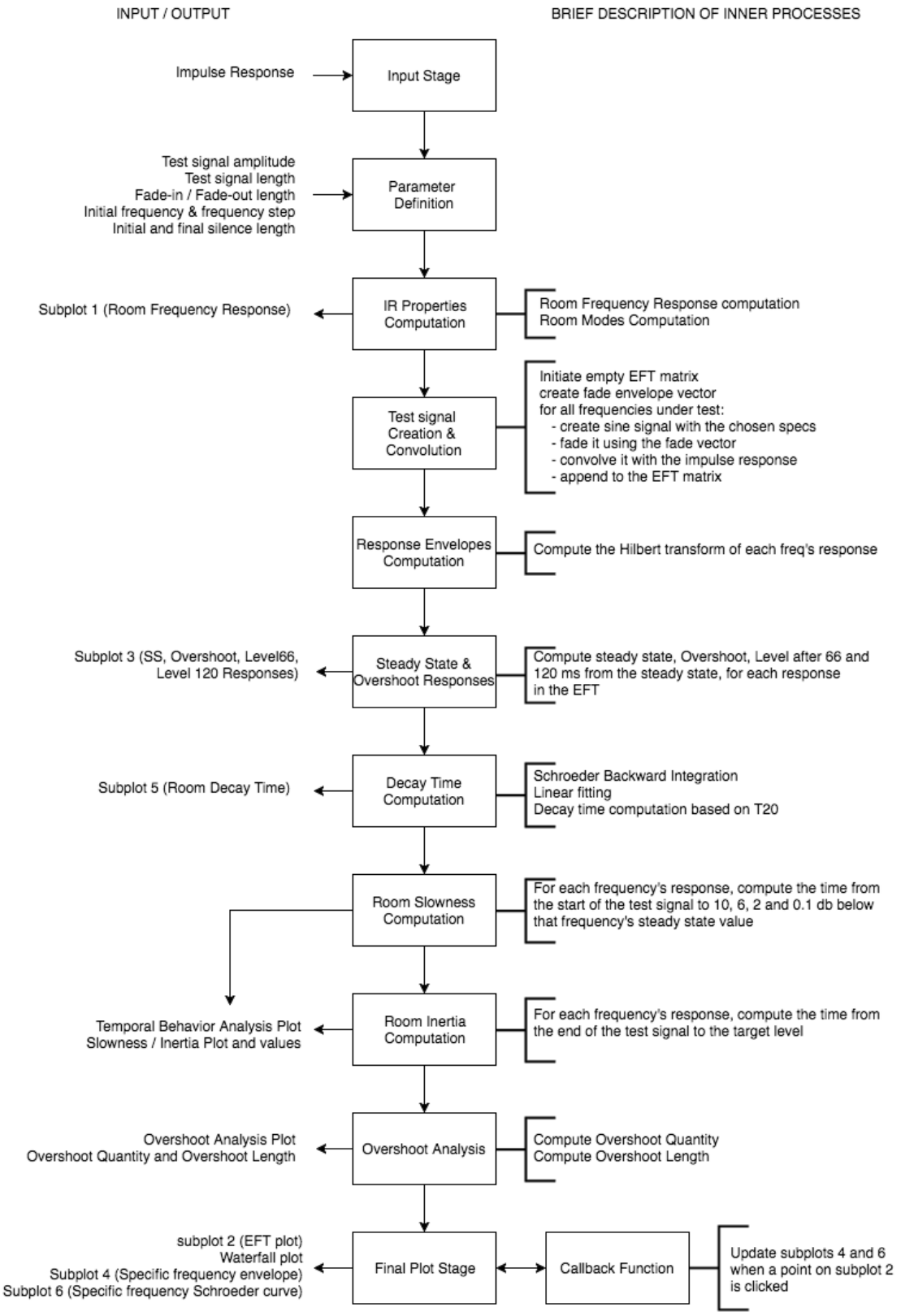


Figure 5.1: AQT Matlab™ Algorithm scheme as developed by the author

All deciBel computation were done by using the command $20log10(abs(fft(quantity)))$ in Matlab™. Since the measured IR is a recorded quantity (acquired with an audio interface) its unit of measurement is Volts, and the deciBel computation is assumed to be performed with respect to 1 Volt. In this context, these deciBels are not dB SPL. What is important, for this research, is the presence of particular behavior in the frequency response, while the actual reproduced level differs by a constant that depends by the output volume in each particular listening condition.

5.2 Room temporal domain analysis

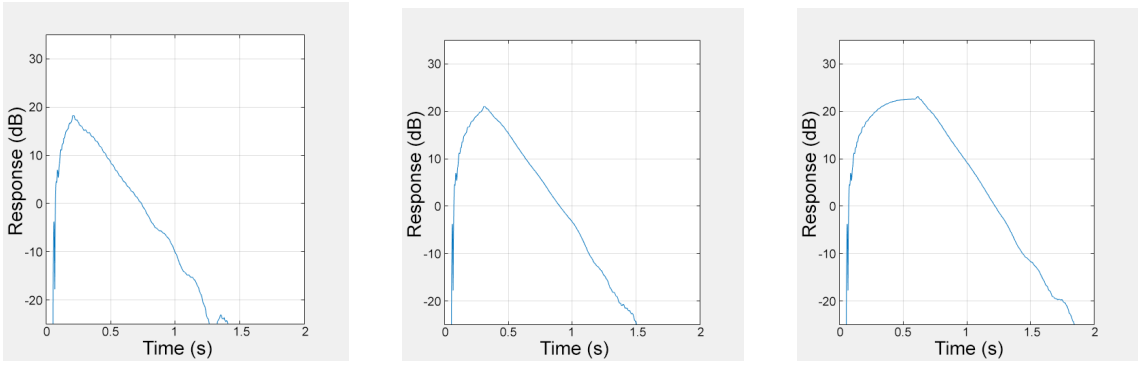
Initially, a simple analysis was performed, computing the system's Steady State Response and the Overshoot Response and inspecting the EFT Plot and AQT Response Envelopes. After the first psychoacoustic test, a further stage was included in the algorithm which performed advanced Overshoot Analysis.

Upon initial inspection of the EFT Plot and energy build-up transients in the AQT Response Envelopes it is possible to see that frequencies develop in time in very different ways.

5.2.1 Behavior on Frequency Response's Peaks

On the peaks of the room's frequency response, the AQT Response Envelopes are usually slow, in time, in reaching their steady state value. In fact, for very short test tones, some frequencies may not reach their final steady state level at all. These frequencies are also the slowest ones to decay, as will be made clear in the following sections. However, not all rooms react equally. Some rooms show very fast envelopes even on the frequency response's peaks.

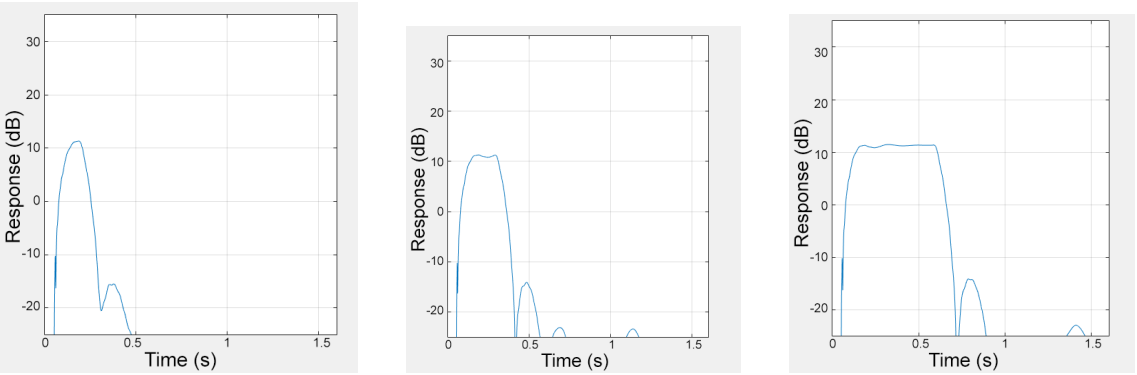
An example of this behavior can be seen in Fig. 5.2, which shows the AQT Response Envelopes for three test bursts at 44 Hz of length 150, 250 and 550 milliseconds in room CN. With a 150 milliseconds exciting tone, the energy level is clearly lower. Only when the burst duration is higher test tones are actually able to reach their stead state value.



(a) 150 ms test tone (b) 250 ms test tone (c) 550 ms test tone

Figure 5.2: Example of slow behavior on Frequency Response's peaks

Fig. 5.3, instead shows the same situation in a room that reaches fastly its steady state value at a similar frequency (Room PRZ, 38 Hz). Even with a 150 milliseconds burst, the steady state value is reached.



(a) 150 ms test tone (b) 250 ms test tone (c) 550 ms test tone

Figure 5.3: Example of fast behavior on Frequency Response's peaks

As a matter of fact, different frequencies behave differently with respect to the speed of rise and decay of the Response Envelopes. These parameters also differ room by room and two new metrics, called "Room Slowness" and "Room Inertia", will be introduced to account for this behavior in the following sections. This behavior is very interesting and some questions have been developed in order to evaluate the psychoacoustic perception of this phenomenon during the tests.

5.2.2 Overshoot Peaks

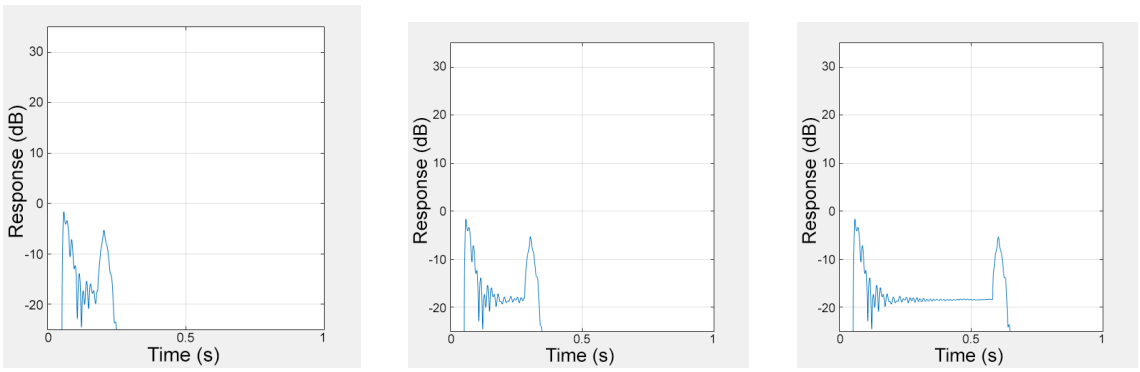
On the valleys of the room's frequency response, and in intermediate points between peaks, two spikes in the AQT Response Envelope can be seen, one at the beginning (opening overshoot) and one at the end (closing overshoot). Sometimes one of the two spikes is lower or almost non-existent, especially the later one. In [8], this transient behavior is associated to the regions were the phase of the direct field is opposite to the phase of the reverberant field. The central region, whose level is the frequency's actual steady state value, is lower than the peaks because of the interference between the two sound fields ([8]).

Initially, "Suono e Vita" hypothized that this behavior could be more psychoacoustically significative with respect to the classic Frequency Response when evaluating the levels of short sounds.

The overshoot behavior (overshoot amplitude and duration) does not change in amplitude with the length of the test signal. Fig. 5.4 shows this behavior, testing the same frequency in the same room with tones of different length (Room PRZ, 110 Hz). The overshoot level, in this case, is about 15 dB higher than the steady-state value. As will be discussed in the following sections, the overshoot amplitude depends, among other factors, also from the initial slope and attack of the test tone and the amplitude of the opening overshoot is generally higher than the closing overshoot. This is expected as a reaction to sharp energy intake.

Part of this research aim is to prove or refuse the hypothesis, presented in [8] that the overshoot response can be more significative than the frequency response when evaluating the perceptual volume of really short sounds.

Figure 5.5 summarizes all the previously described phenomena, showing that some response envelopes on peaks of the FFT have a "slow" behavior, meaning that they do not reach their steady state level for short bursts, while others do ("fast"



(a) 150 ms test tone (b) 250 ms test tone (c) 550 ms test tone

Figure 5.4: Example of Overshoot Behavior

behavior). Response Envelopes on valleys also show Overshoot behavior. This phenomena also recalls the behavior of the step response in higher order systems, in the underdamped and overdamped cases. The link between these two area should be object of further studies (slow reaction is typical of reactive circuits and second order systems).

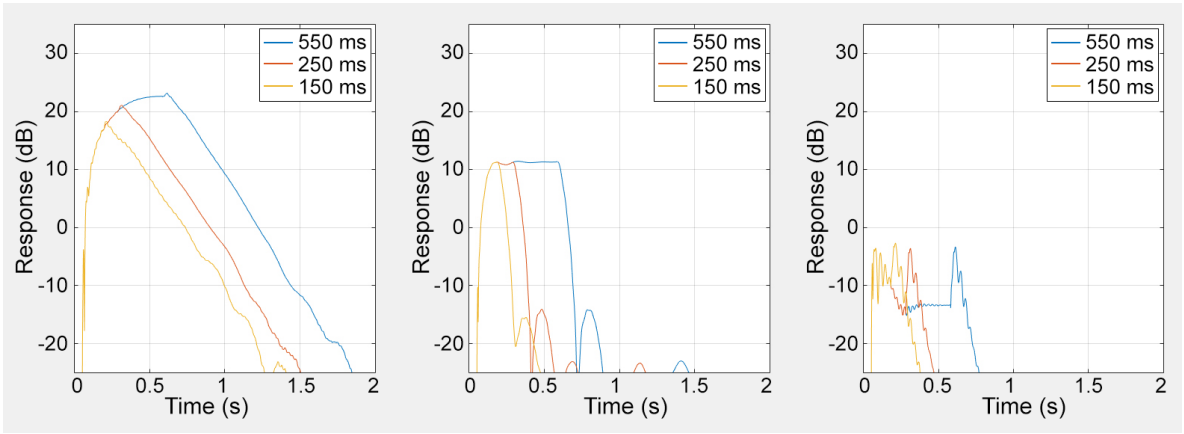


Figure 5.5: Temporal Responses at Low Frequencies in Small Rooms - "Slow" and "Fast" behavior on peaks, Overshoot behavior on valleys

5.2.3 Steady State Response and Overshoot Response

Fig. 5.6 shows the room Steady State and Overshoot Response for room PRZ, both generated with 550 milliseconds bursts.

The Steady State Response is defined, for each frequency, as the value of the AQT Envelope Response at the instant when the AQT test tone stops. This value represents the level that that frequency's AQT Envelope Response reached in that room when its length is the same as the test tone one. It differs from the Frequency Response, since the Frequency Response shows the actual Steady State value for each frequency as though the test tones were infinitely long and were always allowed to reach their steady state. If the test tone lasts long enough, the Steady State Response and Frequency Response are indeed the same curve. For this analysis,

which is focused on the behavior and perception of short sounds, it is important to distinguish between the two curves.

The Overshoot Response is defined, for each frequency, as the maximum value of its AQT response envelope. Fig. 5.6 shows that there are some frequency areas where the Overshoot Response is higher than the Steady State Response. This means that their AQT Response Envelope reached, at some point, higher values (overshoot behavior).

Because of the definition of the Overshoot response, on the peaks of the frequency response this value is usually the same as the steady-state value, since the maximum is the steady state value itself (or the maximum value that that frequency reaches quickly in the test time). On valleys, it is the maximum between the opening and closing overshoots. Intermediate frequencies have intermediate behavior (they may show little overshoot behavior) that depend on the specific room conditions.

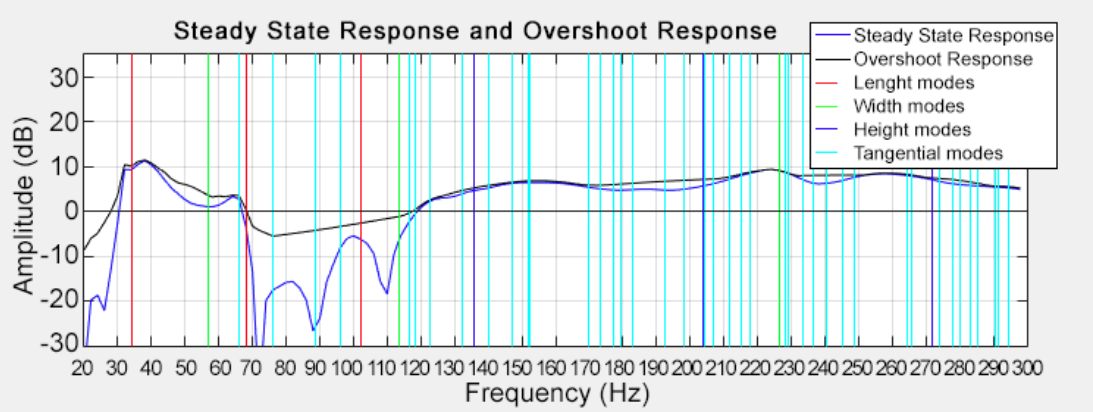


Figure 5.6: Steady State and Overshoot responses (Room PRZ, 550 msec)

Fig. 5.7 shows the same two curves for room CN. The Steady State response features less deep valleys than room PRZ, and Steady State and Overshoot curves are closer with respect to room PRZ.

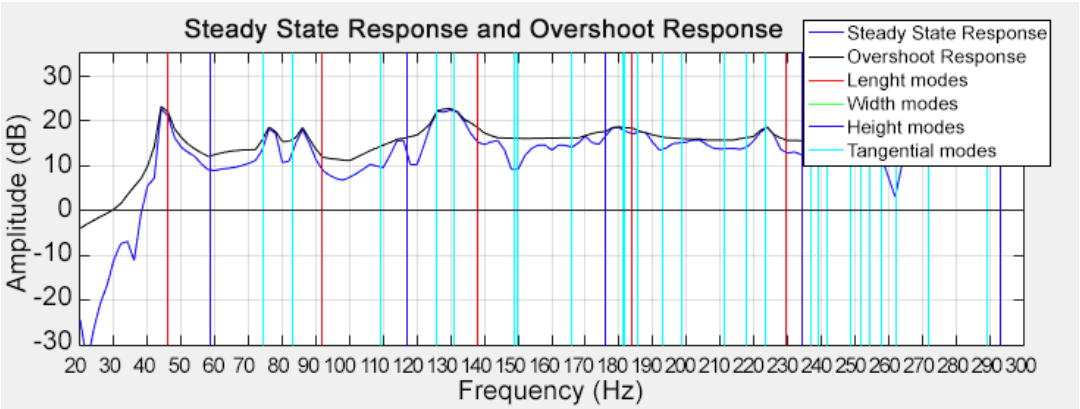


Figure 5.7: Steady State and Overshoot responses (Room CN, 550 msec)

As the rest of the parameters generated by the AQT algorithm, both curves were born starting from a specific burst length duration. However, for some rooms and some frequency response's peaks, the AQT Response Envelope is slow in reaching its steady state value, and for short bursts it may not reach it at all. This, of

course, impacts the Overshoot Response because of the way the response has been defined. In the next sections, after other important parameters and concepts have been introduced, the impact of changing the duration of the test burst over the Overshoot Response will be analyzed.

5.2.4 Advanced Overshoot Analysis

The Advanced Overshoot Analysis block was introduced after the first psychoacoustic test was finished. It computes two parameters for each frequency: the Overshoot Energy Level and the Overshoot Length.

Overshoot Energy Level

The Overshoot Energy Level is defined, for each frequency, as the difference between the Overshoot Response and the Steady State Response. It measures the amount of dB by wich the Overshoot Response exceeds the Steady State Response. Because of the definition of Overshoot Response (for each frequency, the maximum value of its Response Envelope), and of Steady State Response (for each frequency, the value of the Response Envelope at the moment of the burst's ending), the Overshoot Response value is always greater than or equal to the Steady State Response value for all frequencies. This means that, by definition, the Overshoot Energy Level is greater than or equal to zero for all frequencies.

In particular, two variants have been developed: one with an overall Overshoot (computed as the maximum of the AQT Response Envelope for each frequency) and one with an initial Overshoot (computed as the maximum of the first half of the AQT Response Envelope for each frequency). This was aimed at seeing if the initial overshoot was usually the higher between the initial and closing overshoot. Fig. 5.14 shows that this is usually the case.

Fig. 5.14 shows how the Overshoot Energy Level is directly relatable to the presence of a valley in the frequency response. This is confirmed by "Advanced Overshoot Analysis" plots for all rooms, shown in the next sections. This analysis has been repeated for all eight rooms, showing that the Overshoot behavior is indeed present only in valleys and that in all valleys, overshoot behavior is present to some degree.

Overshoot Duration

The Overshoot Duration is aimed at quantifying the duration, in seconds, of the initial overshoot. This quantity has been defined only for those frequencies which shows clear overshoot behavior. In this algorithm, the author decided to define this quantity only for those frequencies where the difference between Overshoot value and Steady State value was greater than 1 dB. For all other frequencies, the Overshoot Duration is set to zero by default. This choice was made because it does not make sense to define this quantity for frequencies that do not show overshoot behavior. Even more so, since the Overshoot Response vector is defined as the maximum of the AQT Response envelope for each frequencies, and for the frequencies that don't have evident overshoot behavior, the maximum is somewhere in the middle of the Response Envelope (sometimes it is equal to the steady state, or a lower value if it's a slow room mode) and the Overshoot Duration would be an incorrect, high value.

This algorithm finds the maximum of the first half of each frequency's AQT Response Envelope, which is the initial overshoot. Then, it looks for the first sample after the overshoot for which the difference between the AQT Response Envelope and that frequency's Steady State value is under a certain threshold (0.5 dB). The algorithm records this point's position in samples and subtracts the number of samples of initial silence in the AQT test signal, returning the duration (converted to seconds) between the start of the AQT test signal and the point, after the overshoot, with value close to the Steady State.

This algorithm does not take into consideration the ripple effect that sometimes happens after the initial overshoot in the AQT Response Envelope, since defining a threshold value that works in all cases might be not possible.

5.3 "Room Slowness" (Opening Transient)

Room Slowness is a new metric aimed at describing the slowness of the room's response at each frequency, which means, how slowly (or fastly) does the room react when excited with a specific frequency.

Room Slowness is computed, for each frequency, as the time between the beginning of the test tone and the moment when the AQT response envelope reaches the steady state value of that specific frequency, minus a threshold value. Four variants have been developed, with 10, 6, 2 and 0.1 dB as threshold values with respect to the steady state level.

The measures with 10 and 6 dB of threshold might be more psychoacoustically significative, since they represent the initial slope of the AQT response envelope. If this time is slow, it means that the frequency gets to its steady state slowly.
The measures with 2 and 0.1 dB of threshold show the last parts of the slope, therefore they might be psychoacoustically non perceptible. In fact, these variants measure the time that the AQT Response Envelope takes to reach a value that is very close to the steady state value. For some frequencies, it happens that there is a fast initial growth, that reaches a value that is slightly lower than the steady state one (but still quite close), while it takes a lot longer to go, in the AQT Response Envelope, from this value to the target level (steady state minus threshold), and this happens with a really low slope. The first part of the AQT Response Envelope (the fastly growing part) is psychoacoustically more meaningful, since it reaches a value that is closer to the steady state one. This aspect will be inspected with a question in the second psychoacoustic test.

As an example, if the lower (10 and 6 dB of threshold) curves feature a very low value, and the higher ones (2 and 0.1 dB) feature a very high value, it means that that frequency fastly goes to a value near its steadystate, and then continues slowly towards its steady state value.
The author's hypothesis is that the lower curves (10 and 6 dB of threshold) are therefore more psychoacoustically important. Fig. 5.13 shows these four curves alongside the "Room Inertia" metric.

In order to describe this parameter with just one number, an overall Room Slowness score has been given to each room. This value was calculated as follows: the Room Slowness variant with 10 dB threshold value with respect to the steady state value was selected. The Room Slowness is a number (measured in seconds) defined for each frequency in the AQT script, which are all frequencies between 20

and 300 Hz. Therefore, in the Matlab™ script, these values were stored in a vector. The global Room Slowness value has been defined as the medium value of this vector, and it represents the medium value of Room Slowness for all frequencies between 30 and 300 Hz. However, the whole curve featuring all Slowness values with respect to frequency still brings important information and has been used to develop focused questions on specific frequencies.

5.4 Room Decay Time

A regular Room Decay Time computation has been performed as part of the AQT algorithm. The script used in this part of the AQT algorithm was written by L. Rizzi and G. Ghelfi from "Suono e Vita". This part of the algorithm excites the impulse response with multiple frequency bursts (slightly different from the AQT signal), computes the Schroeder Backward Integration and computes the decay time using the slope over a fall in amplitude of 25 dB from the steady state value at the end of the tone burst and averages the responses.

The classic Room Decay Time seems to fail to intuitively describe the specific behavior of the closing part of each frequency's Response Envelope. In fact, despite providing a very intuitive plot that shows the decay trend with respect to the frequency axis, the way it is computed makes it hard to understand how fast each frequency is actually decaying. Since both the starting amplitude (steady state value) and the target amplitude (25 db under the starting amplitude) depend on the specific frequency and on its steady state value, if the steady state value is too low (as an example, in the valleys of the frequency response), it may happen that the target amplitude is at the end of the Schroeder Backward Integration curve. This returns a very high value for that frequency's decay time, even though that specific frequency's AQT Response Envelope actually decays very fastly. This is not completely representative of the slope of that frequency's AQT Response Envelope, therefore a new metric was developed, called "Room Inertia".

In order to describe this parameter with just one number, an overall Room Decay Time score has been given to each room. This value is the medium value of the vector containing the Decay Time values for each frequency between 30 and 300 Hz.

5.5 "Room Inertia" (Closing Transient)

Room Inertia is an alternative metric to the Room Decay Time, aimed at solving the aforementioned problem.

Room Inertia is computed as the time between the end of the tone burst and the moment when the AQT response envelope reaches a fixed, "target" value. The target value is arbitrarily defined as the minimum value in the Frequency Response minus 1 dB. This way, the target value is the same for all the frequencies (instead of depending on each frequency's steady state value as in the Decay Time) and is always greater than zero. The target value definition is arbitrary and may be changed, since what is interesting is not the Inertia value per se, but the trend in the frequency domain and the connection between high "Room Slowness" and "Room Inertia" values. Since the target value is always the same, this metric is more significative of the actual slope of the AQT Response Envelope in the closure

part, and it allows for an easier visualization of the time that each frequency needs to decay from their steady state, with respect to the Decay Time.

As it is possible to see in the next sections, the Temporal behavior analysis plot of each room shows that there is a direct correlation between the frequencies that have high Slowness and high Inertia, and that this happens mainly on the peaks of the Frequency Response.

In order to describe this parameter with just one number, an overall Room Inertia score has been given to each room. This value is the medium value of the vector containing the Room Inertia values (exactly like it has been done for the Room Slowness single Room value) for each frequency in the domain between 30 and 300 Hz.

Since each room has finally been described with a single value for each of the parameters (Room Slowness, Room Decay Time, Room Inertia), the rooms were ranked with reference to these values, and these ranks were taken into account while preparing the second psychoacoustic test.

5.6 Correction of Artifacts in some Room Inertia and Decay Time Responses

Some Impulse Responses have lead to the generation of small artifacts in the Room Inertia and Decay Time Responses. In particular, the artifacts were spikes in the Room Inertia value and Decay Time values in correspondance with some frequencies where the frequency response presented a very sharp change, almost like a discontinuity, even if very small. An example of such artifacts can be seen in Fig. 5.8 at 120 Hz, where the frequency response has a little discontinuity and the Inertia Response has a sharp peak.

It is not completely clear what caused such artifacts in the frequency responses, even though the authors think this could be caused by little problems happened during the measurements of the impulse responses. Further research could be aimed at solving this ambiguity. Two of the eight rooms used in this research presented some minor artifacts in their Room Inertia responses. These rooms were PRZ and SNT, and four of the eight rooms presented such artifacts in their Decay Time responses: room DrmA, DrmB, SGR, SNT.

This problem was taken care of, in both Inertia and Decay Time, by inserting a condition that tested all elements of those vectors looking for "outliers" (elements that were really far from the neighbors). This cycle tested if each element was greater than the previous and next elements of the vector by at least a certain amount (which was set at 0.5 seconds for Inertia and 2 seconds for Decay Time). For elements that verified this condition (meaning that they were "outliers", the element was substituted with the average value of the previous and next elements in the same vector. Fig. 5.8 shows the original Inertia response for Room PRZ with an artifact at 120 Hz. Fig. 5.9 shows instead the same response after the aforementioned correction.

The Inertia and Decay Time Responses in the following chapters are shown with correction. In particular, rooms PRZ and SNT feature Inertia Response correction and rooms DrmA, DrmB, SGR, SNT feature Decay Time Response correction.

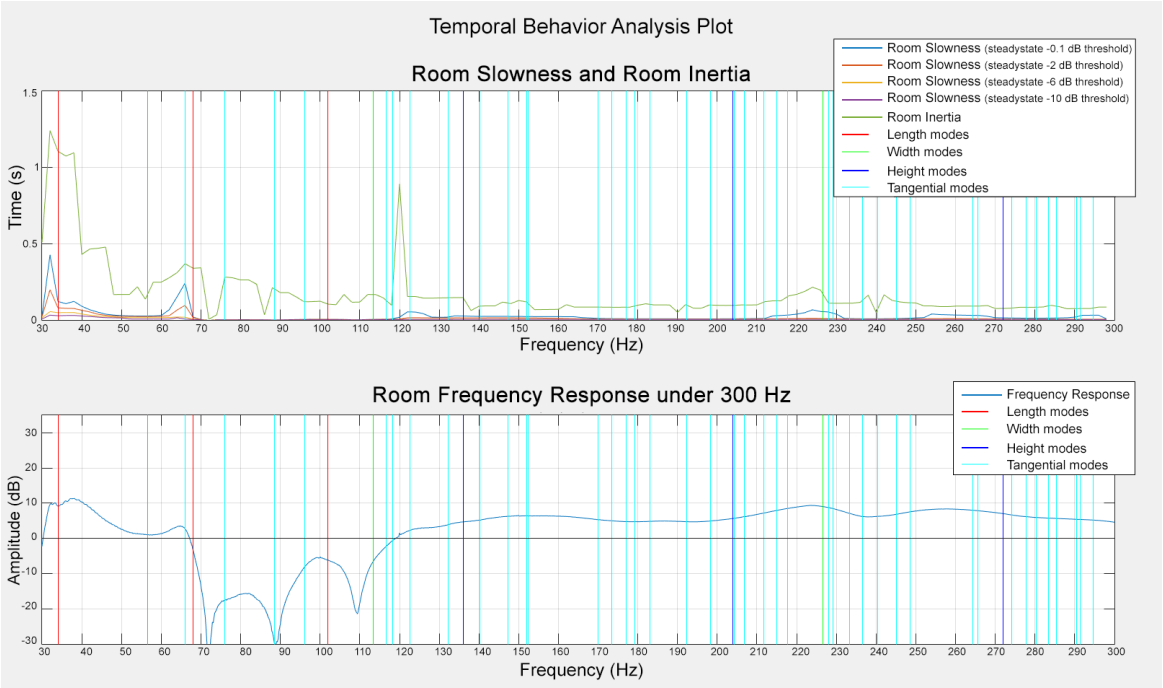


Figure 5.8: Room PRZ - Artifact example in the Inertia Response - Before Correction

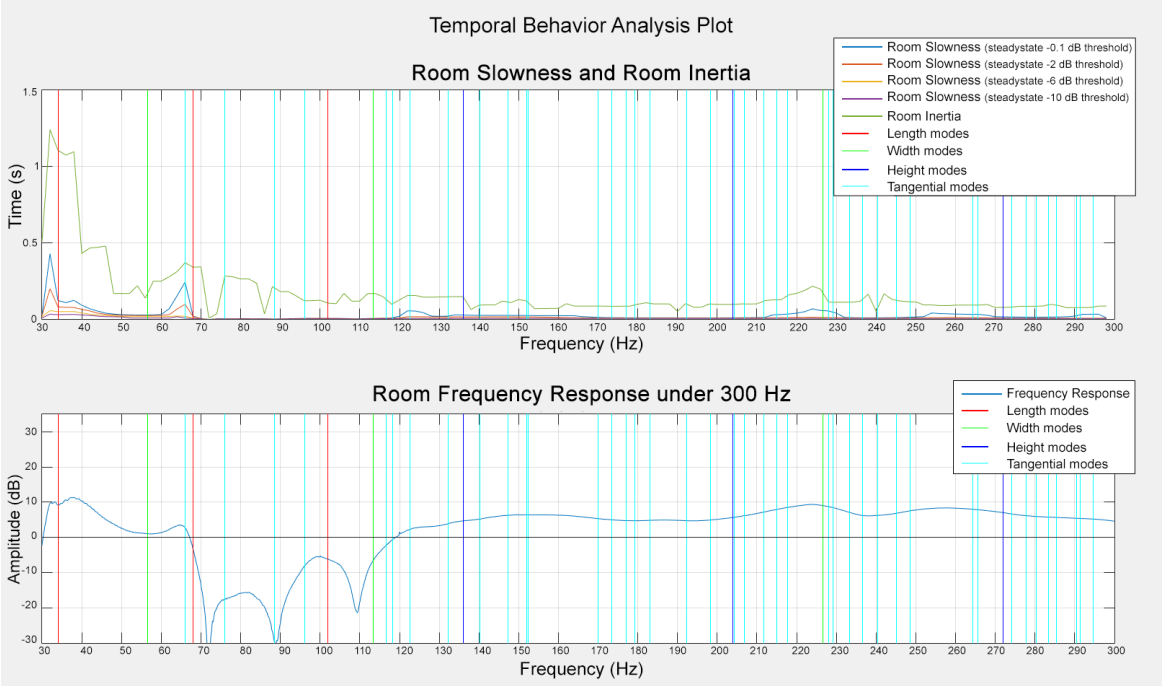


Figure 5.9: Room PRZ - Artifact example in the Inertia Response - After Correction

5.7 AQT Algorithm Output

The AQT algorithm results are shown through four figures:

1. Waterfall Plot
2. Main AQT Plot
    - Room Frequency Response
    - Steady State Response, Overshoot Response (originally also Level after 66 ms, Level after 120 ms)
    - Room Decay Time
    - EFT (Energy / Frequency / Time) Plot
    - Specific frequency AQT Response Envelope
    - Specific frequency Schroeder backward integration curve
3. Temporal Behavior Analysis
    - "Room Slowness" and "Room Inertia"
    - Room Frequency Response
4. Overshoot Analysis
    - Room Frequency Response and Overshoot Energy Level
    - Overshoot Duration

All plots whose x-axis is frequency also feature vertical lines that represents the room axial and tangential modes. This addition was suggested by "Suono e Vita" in order to enrich the room analysis and description. Axial modes are depicted with different colors in order to distinguish which modes were relative to a specific room dimension, while tangential modes are all light-blue, as they are less important for our analysis. These room modes were calculated with eq. 2.15, which is the classic formula for the shoe-box room model. Therefore, the computed modes may be moved from frequency peaks and dips, when the real rooms were not always a perfect parallelepiped. When this happens, the frequency response curve has priority over the computed room modes.

5.7.1 Waterfall Plot

The waterfall plot shows the amplitude, in dB, of each AQT Response Envelope from frequencies ranging from 20 to 300 Hz, with respect to time. In practice, each section alongside the frequency axis (parallel to the time axis) is the AQT Response Envelope of that specific frequency.

Fig. 5.10 shows the waterfall 3D Plot from two different points of view for room PRZ. It is possible to see that modal frequencies at the very low end of the spectrum have a Response Envelope which is slower in arriving to the steady state value and slower in decaying, while, for upper frequencies (even frequency response's peaks), this behavior is less evident. Room PRZ is actually one of the rooms with

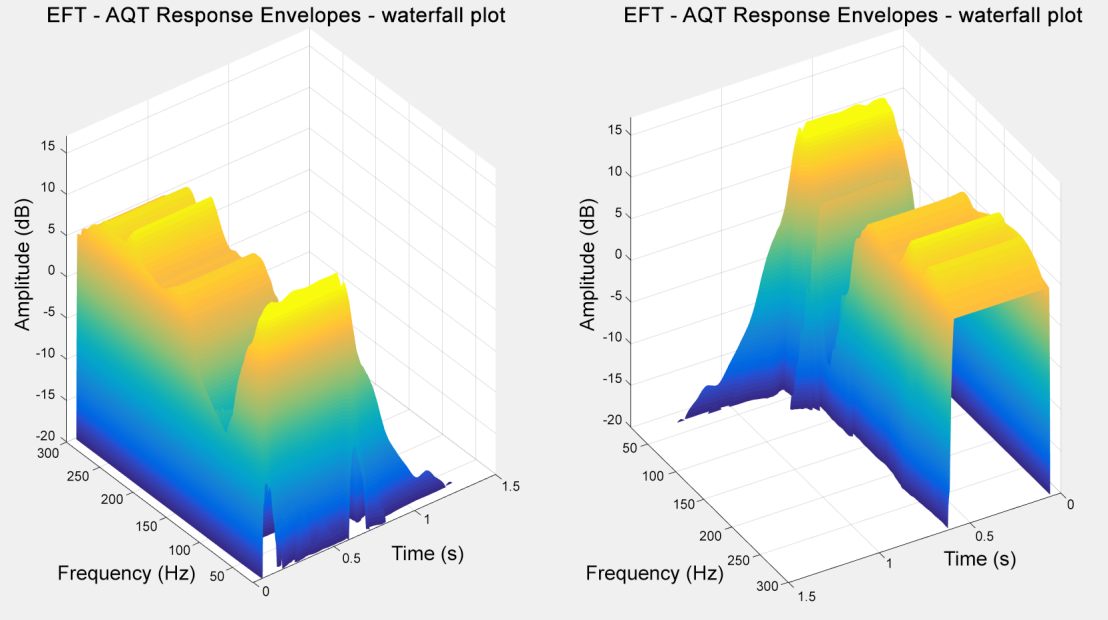


Figure 5.10: Waterfall 3D Plot (Room PRZ)

slowest Inertia times, and this is why most frequencies' energy level decays so fastly. "Slower" rooms will have longer decay tails even at higher frequencies.

Fig. 5.11 still shows the same waterfall 3D plot, slightly rotated. This allows to see that, on the huge frequency response valley (and, to a smaller extent, also on the valleys that are higher in the frequency domain) there is an higher amplitude value near the beginning of the x-axis. This is, of course, the opening overshoot behavior.

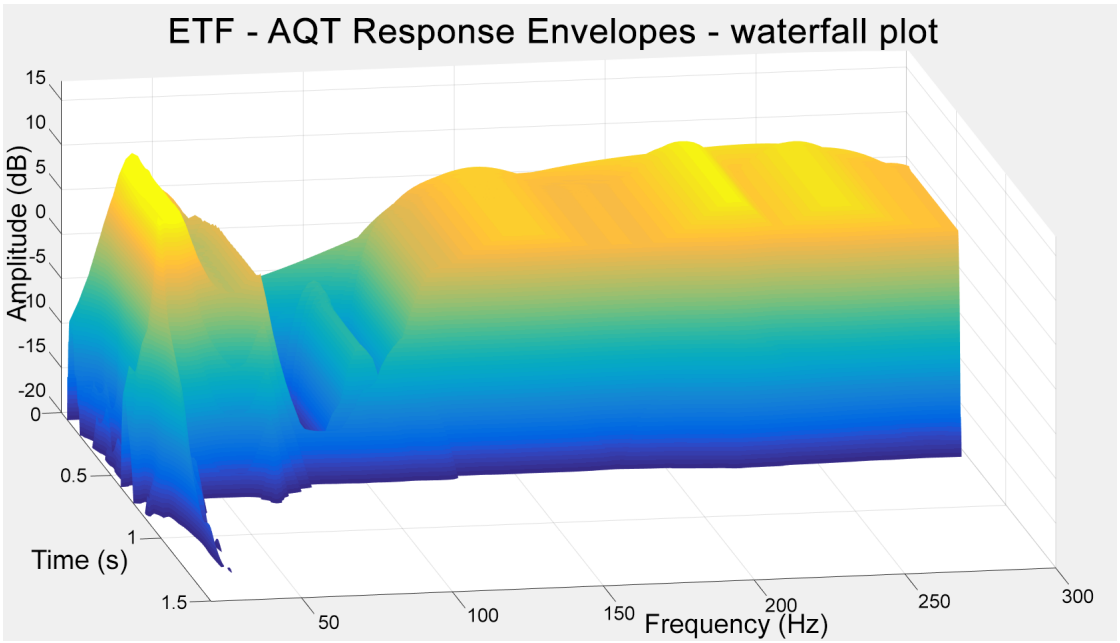


Figure 5.11: Waterfall 3D Plot (Room PRZ), different angle

These waterfall plots do not provide further information, as they are only a different view of the EFT Plot that will be introduced shortly. However, these plots allow to see more clearly the shape and temporal behavior of AQT Response

Envelopes.

5.7.2 Main AQT Plot

Fig. 5.12 shows the main output plot of the AQT algorithm, featuring Room Frequency Response, Steady State and Overshoot Response, Room Decay Time, EFT Plot, AQT Response Envelopes and Schroeder backward integration curves.

The plots "AQT Response Envelope" and "backward Schroeder integration curve" appeared for each specific frequency when clicking a point in the EFT subplot, and they show the temporal evolution and Schroeder curve of the frequency that has been clicked on. This was implemented through a callback function on the EFT subplot. A sample image of the output is shown in Fig. 5.12.
The original AQT algorithm did not feature Room Decay and Schroeder integration curve plots. Also, the second subplot originally featured two curves which showed the levels after 33 and 66 ms from the steady state. In [8] the 33ms curve was chosen since it was significant in car systems, which are a lot smaller than rooms. Therefore, 66 and an even higher value which is specific for rooms, 120 ms, have been chosen for this algorithm. While providing an intuitive idea of the levels after some time from the steady state, these curves have not been used since more specific temporal analysis has been performed in other sections, therefore they are not present in the final version of the AQT results.

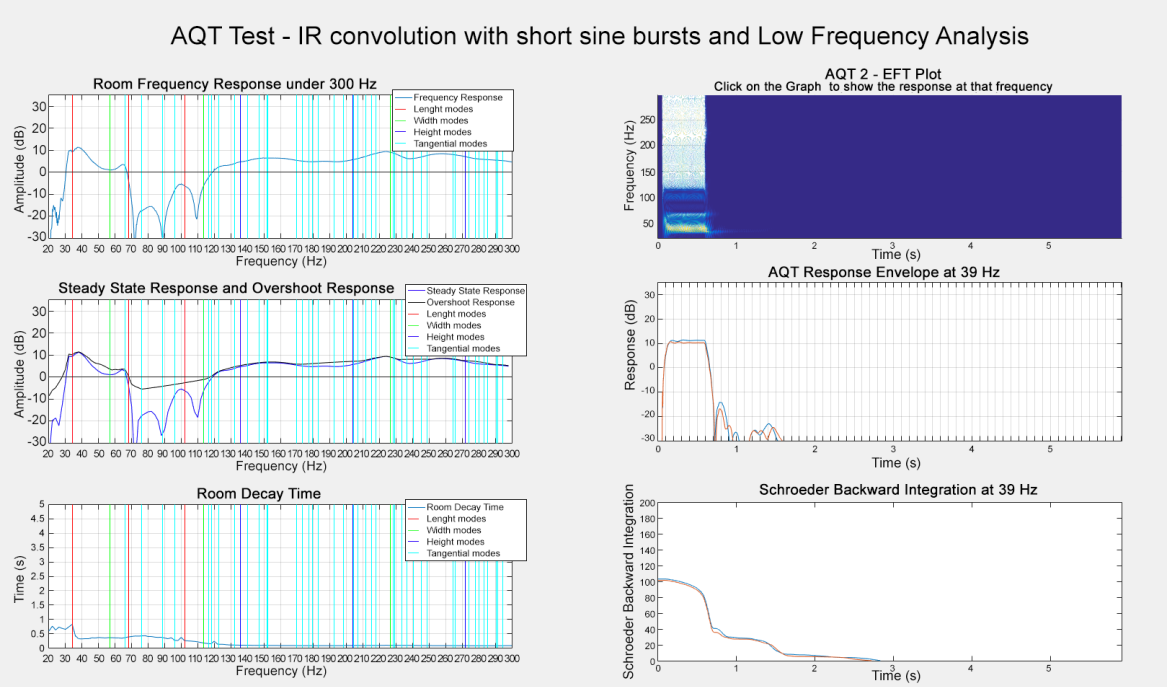


Figure 5.12: Main AQT Plot (Room PRZ)

EFT Plot

The EFT Plot is the second subplot in the Main AQT Plot, and is the primary inspection tool of the AQT script, since it allows to visualize the AQT Response Envelope of each frequency. The EFT Plot features time on its x axis, frequency on its y axis, and the color value represents the value of the envelope of each response: blue color when the amplitude is low, white when it is high. This allows to rapidly see which frequencies have higher peaks or longer tails in their response, as well as giving a visual confirmation of what can be already seen by the frequency response, i.e., peaks and valleys. The EFT Plot is the 2D counterpart of the Waterfall Plot introduced earlier.

5.7.3 Temporal Behavior Analysis Plot

The Temporal Behavior Analysis plot features two subplots. The first one shows the four variants of the Room Slowness parameter and the Room Inertia parameter. The second one shows the Room Frequency Response. Fig. 5.13 shows this plot for Room PRZ.

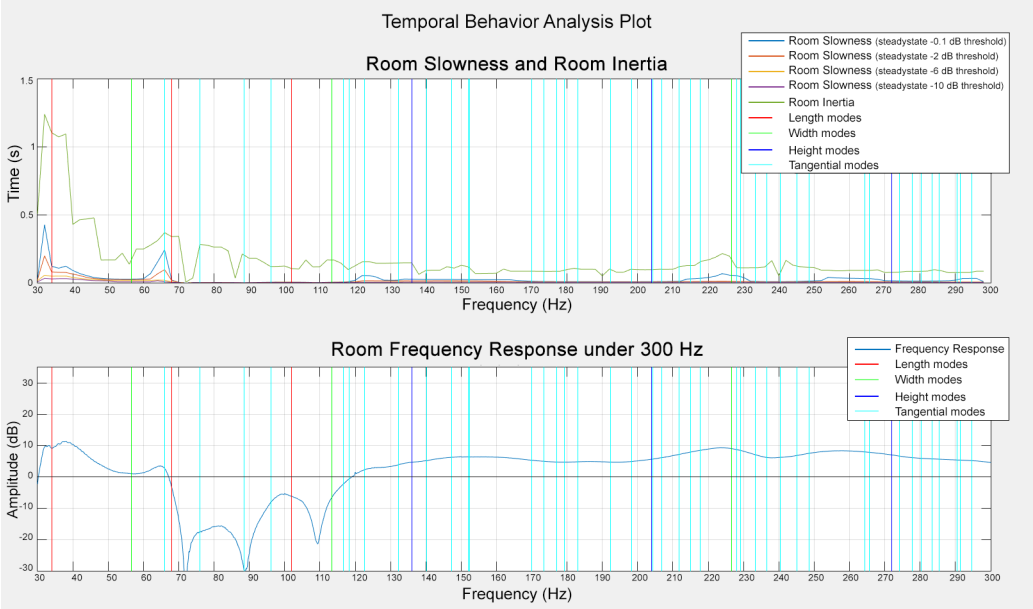


Figure 5.13: Temporal Behavior Analysis Plot (Room PRZ)

5.7.4 Overshoot Analysis Plot

The first subplot features the Room Frequency Response (blue line) and the Opening Overshoot Energy Level (yellow line) defined as the difference between the steady state value and the opening overshoot value for each frequency. The dark orange line is the Overshoot Energy Level, which is the same but computed using the overshoot value (on the whole AQT Response Envelope) instead of the opening overshoot value (on the first half of the AQT Response Envelope). This line is almost always hidden by the yellow one, meaning that the opening overshoot is almost always the higher one. This is expected because the chosen fade-in value is shorter than the fade-out one, producing generally higher overshoots (the choice and test regarding the effect

of fade-in and fade-out values is discussed in the next sections). The second subplot shows the duration (in seconds) of the overshoots, ignoring ripple effects in the AQT Response Envelope. This value is defined only for frequencies which shows evident overshoot behavior, while it is set to zero for the ones who do not, to avoid confusion when reading results. Fig. 5.14 shows this plot for Room PRZ.

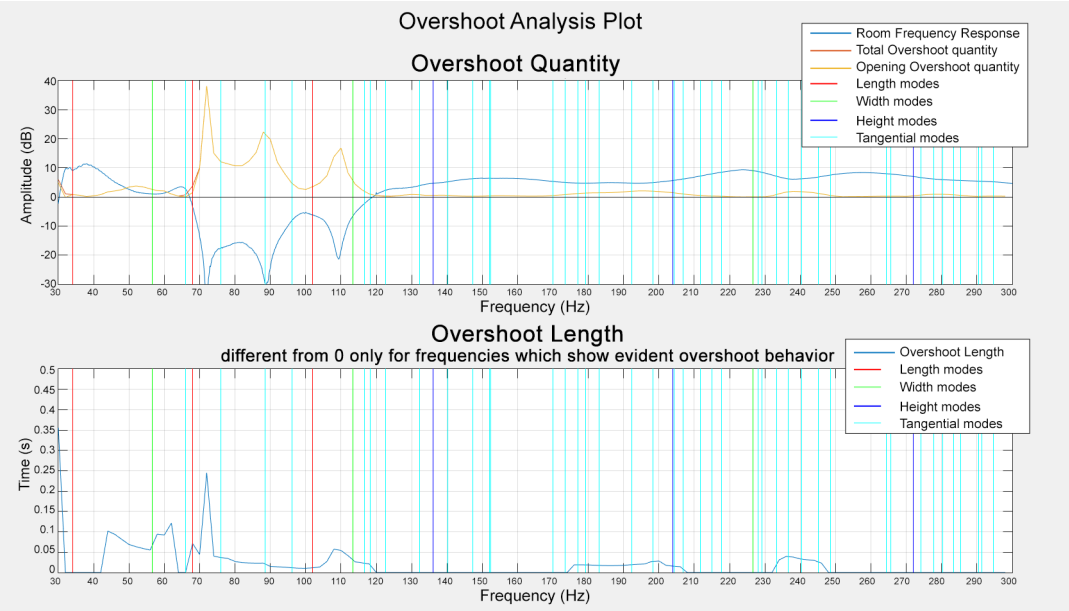


Figure 5.14: Overshoot Analysis Plot (Room PRZ)

5.8 Variability of Overshoot Response with respect to the test burst length

Since the hypothesis of the author, also supported by "Suono e Vita", is that the Overshoot Response might be more psychoacoustically important than the Frequency Response when critically evaluating short sounds, it is important to study what changes in the Overshoot Response and Overshoot Energy Level when changing the duration of the AQT test tone. This has been studied for two different rooms, namely, PRZ and CN which are rooms that show opposite behaviors: PRZ features low "Slowness" and "Inertia" values, and it lets most of its frequencies' AQT Response Envelope reach their steady-state value even for short, 150 milliseconds bursts. CN, on the other sides, has high "Slowness" and "Inertia" values, and some of its frequencies feature EFT Response Envelopes that do not reach their steady-state with short bursts.

The reason behind this analysis arises from the definition of Overshoot Response (once again: the maximum value, in dB, of the AQT Response Envelope for each frequency. That it, the maximum overshoot value where overshoot behavior is present, and the value at the steady state instant when no overshoot behavior is present). In fact, with rooms that feature high values of Slowness and Inertia, some of the peaks of the frequency response (as an example, see Fig. 5.2) do not reach their theorical steady state value (the one shown by the frequency response). Consequently, for lower test tones, the Overshoot Response changes on the frequencies whose AQT

Response Envelope is slower in reaching its steady state value, and this happens in rooms with high Slowness and Inertia values. Rooms with lower Slowness and Inertia values should still show this effect, however in a minor way since their frequency response peaks have AQT Response Envelopes that are able to reach their theoretical steady state value even with short bursts.

Fig. 5.15 shows three different sets of Steady State Response and Overshoot Response curves for PRZ room (which can be considered a "fast" room, regarding Room Slowness and Inertia parameters). The higher subplot shows these curves as they were generated with 150 milliseconds bursts, the middle one with 250 milliseconds bursts and the lowest one with 550 milliseconds bursts.

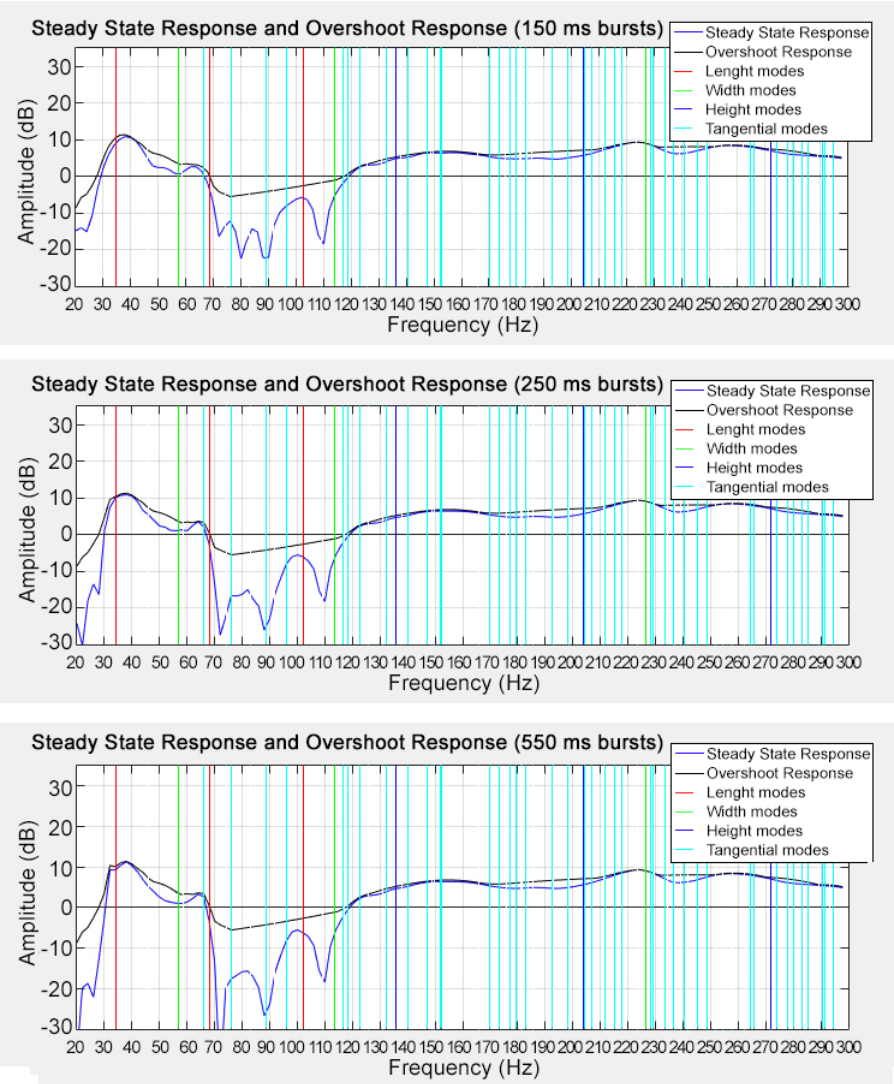


Figure 5.15: Overshoot Response Change with respect to burst duration - Room PRZ

Some difference in the low frequency area is visible, particularly at 32 Hz, where the shortest bursts do not reach their actual steady state (therefore the overshoot response is, by definition, lower) and at 76 Hz, where the shortest bursts have an higher steady state value. This happens probably because of the rippling effects between the very high opening and closing overshoots at that frequency. For longer bursts, at the "steady state instant" (when the test tone stops) the rippling effect of the overshoot is long gone. On the valleys, however, the Overshoot Response is the same for all durations, as expected, since it measures the value of the overshoots which do not change with the sound duration, as already stated. As expected, there is not much change: in fact, this room has low Slowness values, making possible for most of the frequency response's peaks to have an AQT Response Envelope that reaches steady state even with short bursts.

Figure 5.16: Overshoot Energy Level Change with respect to burst duration - Room PRZ

Fig. 5.16 shows the Overshoot Energy Level in the same cases (higher subplot shows curves generated with 150 msec bursts, middle subplot shows curves generated with 250 msec bursts, lower subplot shows curves generated with 550 msec bursts). It confirms what was previously said: the overshoot quantity remains basically the same, besides the points where the steady state value changes for short bursts (as an example, the valley at 76 Hz). It is important to remember, however, that this

does not change the Overshoot Response, since, on valleys, its value is the level of the maximum overshoot.

Fig. 5.17 shows the same analysis of Fig. 5.15 with room CN, which instead has higher values of Slowness, making it impossible for some frequencies' AQT Response Envelope to reach their steady state value for really short bursts.

As expected, more difference is visible in this case with respect to room PRZ in Fig. 5.15. While the change is again limited to the low frequency area, it is more evident. For shorter bursts (first subplot) the steady state and overshoot values are lower also on frequency response's peaks. On peaks, the Overshoot Response is indeed modified, since it depends on the maximum value reached by the EFT Response Envelope.

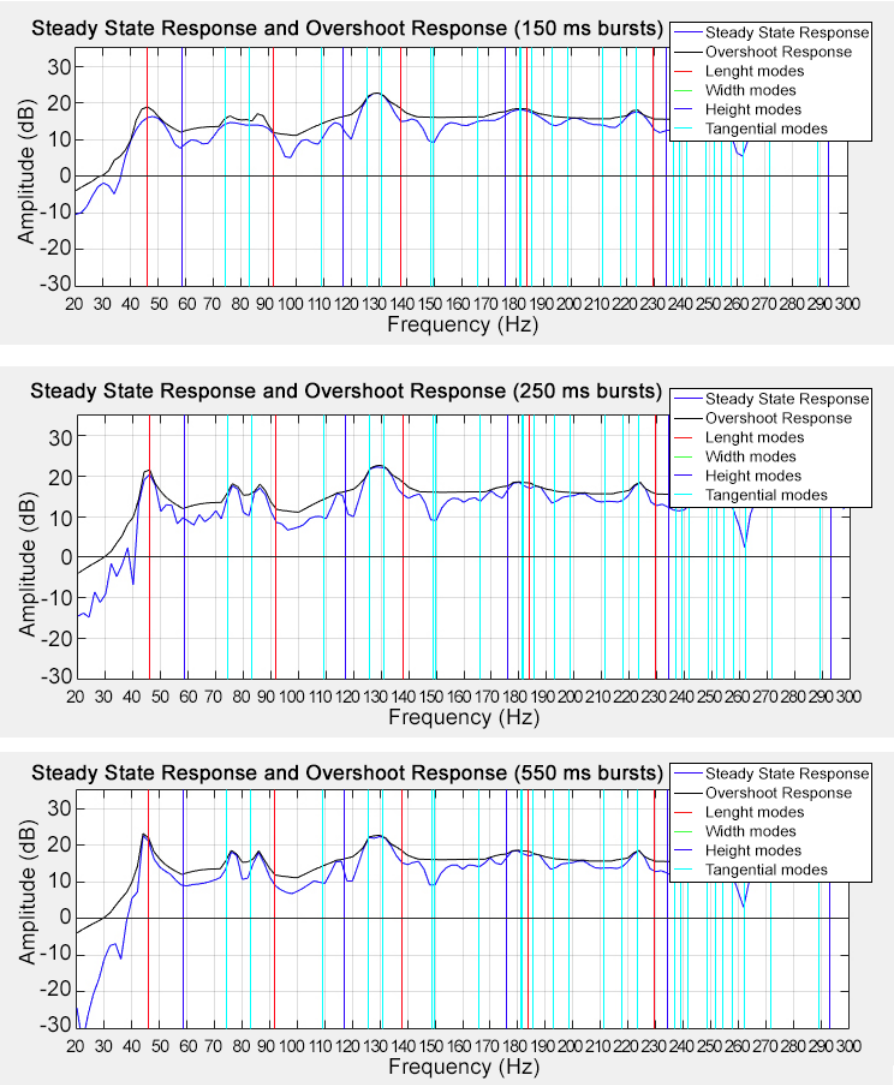


Figure 5.17: Overshoot Response Change with respect to burst duration - Room CN

Fig. 5.18 shows the Overshoot Energy Level in the same cases (again, higher subplot corresponds to 150 msec bursts, middle subplot corresponds to 250 msec bursts, lower subplot corresponds to 550 msec bursts).

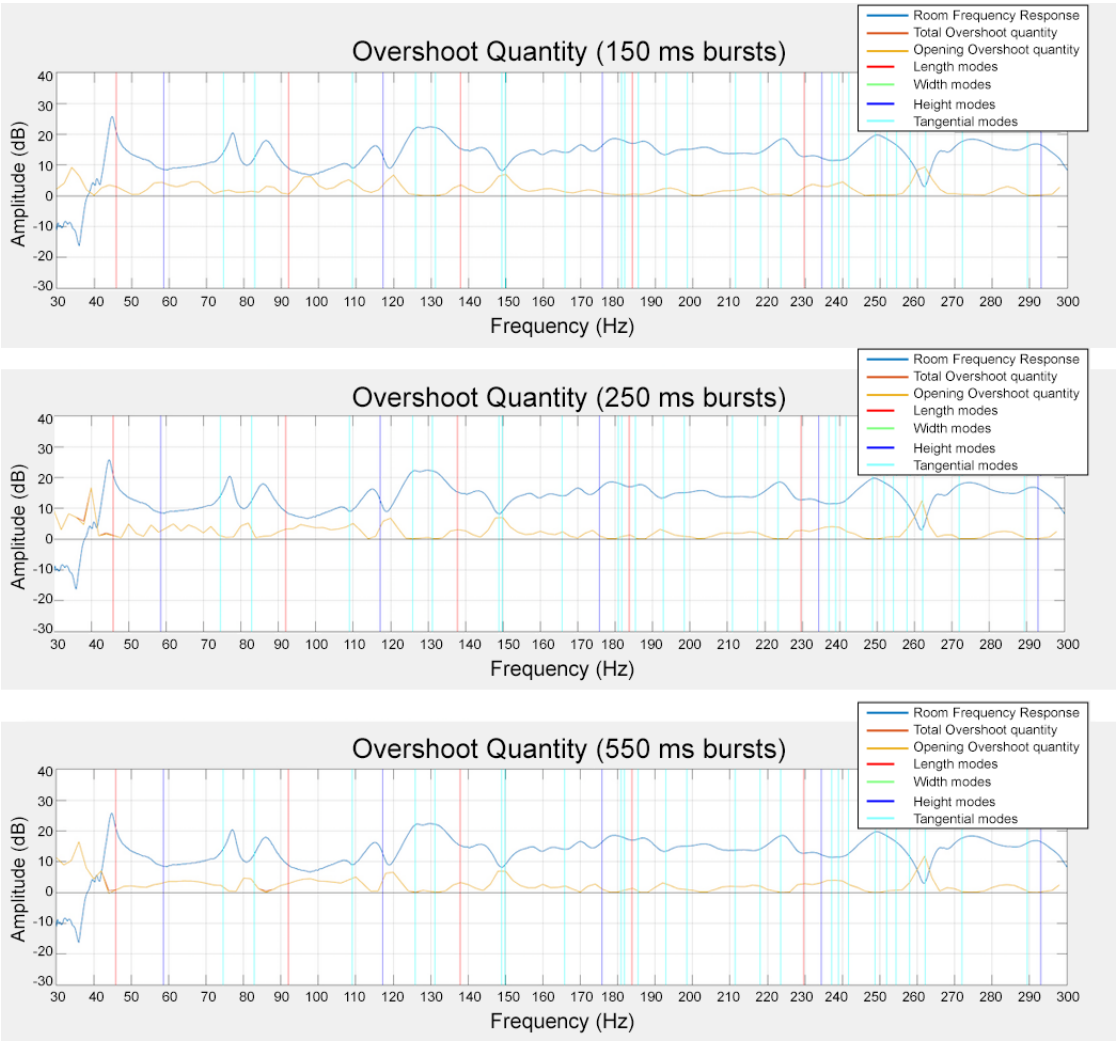
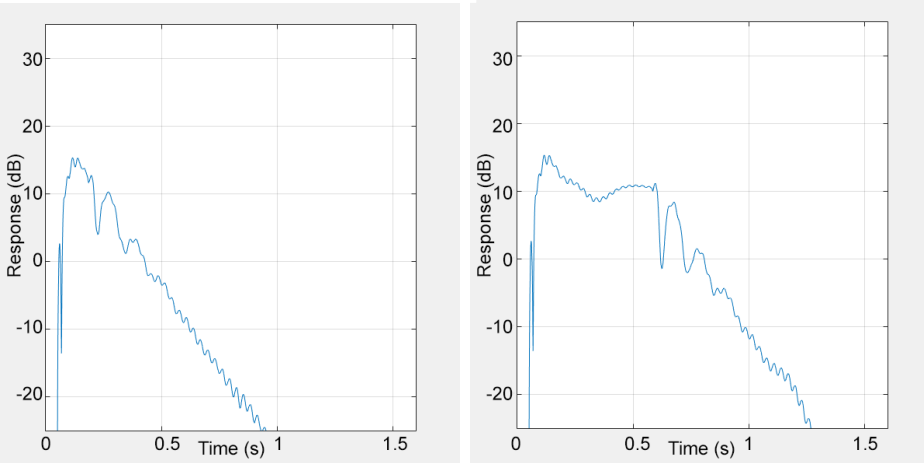


Figure 5.18: Overshoot Energy Level Change with respect to burst duration - Room CN

A particular behavior happens at 80 Hz in room CN: for 150 msec bursts, the Overshoot Energy Level is low, which means that overshoot and steady state are very close. For longer bursts (250 and 550 msec), the Overshoot Energy Level is a bit higher, indicating that the overshoot is higher than the steady state. The corresponding AQT Response Envelopes are shown in Fig. 5.19. This is a very particular case: with the shorter burst (Fig. 5.19a), at the steady state instant (200 milliseconds, which is the burst length plus the initial silence) there is a very tiny peak in the Response Envelope, which makes this value close to the maximum (overshoot value). With the longer burst (Fig. 5.19b), the steady state stabilizes at a slightly lower value at the steady state instant (600 milliseconds), allowing the Overshoot Energy Level to be, by definition, a little higher. This is just a particular case that shows how a little variation in the AQT Response Envelope can cause strange behavior in the output, leading to difficult interpretation of data.



(a) AQT Response Envelope - Room CN, 80 Hz, 150 ms burst

(b) AQT Response Envelope - Room CN, 80 Hz, 550 ms burst

Figure 5.19: Particular Behavior on Overshoot Energy Level in Room CN when varying burst length

This analysis was performed after the two psychoacoustic tests, so it was not possible to test if two tones at the same fundamental frequency on the lowest and slowest room mode with different length in the same room with high Slowness values would be perceived with a different volume. This could be, however, an interesting research for the future.

Of course, the most significative Response Envelope curve for evaluating short sounds should be the one developed by the algorithm with test tones with the exact same short duration as the sounds they aim to evaluate.

However, the most significative could be the Overshoot Response curve developed with 150 milliseconds bursts. This is because, for "faster" rooms (lower Slowness and Inertia values) almost nothing changes with respect to the curve developed with 550 milliseconds bursts, while for "slower" rooms (higher Slowness and Inertia values) the curve developed with 150 milliseconds bursts also captures the lower amplitude value on the frequency response's peaks when compared to the theoretical frequency response steady state values, therefore including also the effect of the room Slowness in the opening transient.

5.9 Choice of adequate test signals

As already stated, the AQT tests the room spectral and temporal response with sine wave pure tone bursts. It is therefore really important to choose a correct duration of such bursts, in order to produce results which are perceptually meaningful. Since this research is aimed towards the perceptual side of acoustics, the choice of test signal duration has been made to test different but realistic conditions, namely, to simulate notes that were long enough to reach their steady-state value and to simulate short notes that didn't reach the steady state, in order to study the difference in perception. Two types of fade-in and fade-out slopes have also been tested, as well as specific fade-in and fade-out values.

5.9.1 Choice of duration - the "real" sounds

First of all, 0.55 seconds was chosen as an initial reference value. This value has been defined in (Citazione Rizzi-Nastasi) and it was suggested as a good value for evaluating the frequency response, since in that case study it was enough for all frequencies to reach their steady state. This duration is really close to 0.50 seconds, which is the duration of quarter notes at 120 bpm (moderato or allegro).
Initially, a conservative value of 1.15 seconds was chosen as the upper limit, being more than double the 0.55 value. However, as hinted in [35], the 0.55 seconds value is enough for most frequencies to reach their steady state value. For the sake of simplicity in the psychoacoustic tests, the 0.55 seconds value was finally kept as the upper limit for the burst duration.

As an intermediate value, 0.25 seconds was chosen. This value has been proposed by Farina ([8]) as the default value for the original AQT algorithm, and, as resulting from the analysis of the previous paragraphs, it is enough for some frequencies to reach steady state, while being often too short for other frequencies to do so.

As the lower limit, 0.15 seconds was chosen. This value has been chosen by the author by averaging some common duration of short notes, like 16th notes at 120 bpm (0.125 seconds), 8th notes at 170 bpm (0.176 seconds). The author felt that this value was significative of fast-paced, low frequency driven modern music. This value allows only a few frequencies to reach their steady state, and it was originally meant to be used for the first psychoacoustic test. However, it was used also in the second test since testers mentioned it gave significant results in the first test.

In conclusions, AQT analyses have been carried out with bursts of 0.15, 0.25 and 0.55 seconds.

5.9.2 Choice of fade-in and fade-out values

Since this work has the aim of describing the psychoacoustic perception of the effects under tests, real sounds have to be used. Obviously, in nature real sounds do not have an infinite initial slope (like a step signal). This confirms the need of using fade-in and fade-out values of the AQT test tones which are different from zero.

The fade-in and fade-out values of the sine bursts are really important, as they can impact the response in the simulation and the interpretation of results. For a correct visualization and interpretation of the analysis results, each sine wave signal envelope needs to be shaped in a way that is similar to the envelope of the real-world, complex sound that the test aims to simulate. As an example, if the AQT algorithm has to simulate the room response to a kick sound, each sine wave should have a short fade-in and a really long fade-out in order to be as similar to a real kick envelope as possible. If this wasn't the case and the fade values were not controlled, the algorithm would produce opening and closing transients in the response envelope that would not be present with real sounds that always have a different envelope.

Fade duration as a percentage of the burst's length

Initially, the author's AQT implementation featured fade-in and fade-out lengths which were 3 percent of the burst duration, since the original AQT definition (Citazione AQT) didn't specify a value. It has soon become apparent that this

was not a correct choice, since shorter bursts featured shorter fade-in and fade-out values, which produce higher overshoot values.

As shown in Fig. 5.20, longer bursts indeed produce lower overshoots. This is also visible in Fig. 5.21. Fig. 5.20 shows the Room Frequency Response (upper plot), Overshoot level at each frequency with different burst lengths (middle plot) and Steady State level with different burst lengths (lower plot). In the second subplot, it is possible to see that longer bursts (purple and green line) have lower overshoot behavior. Indeed, the fade-in and fade-out values are longer, since they are a percentage of the burst's length.

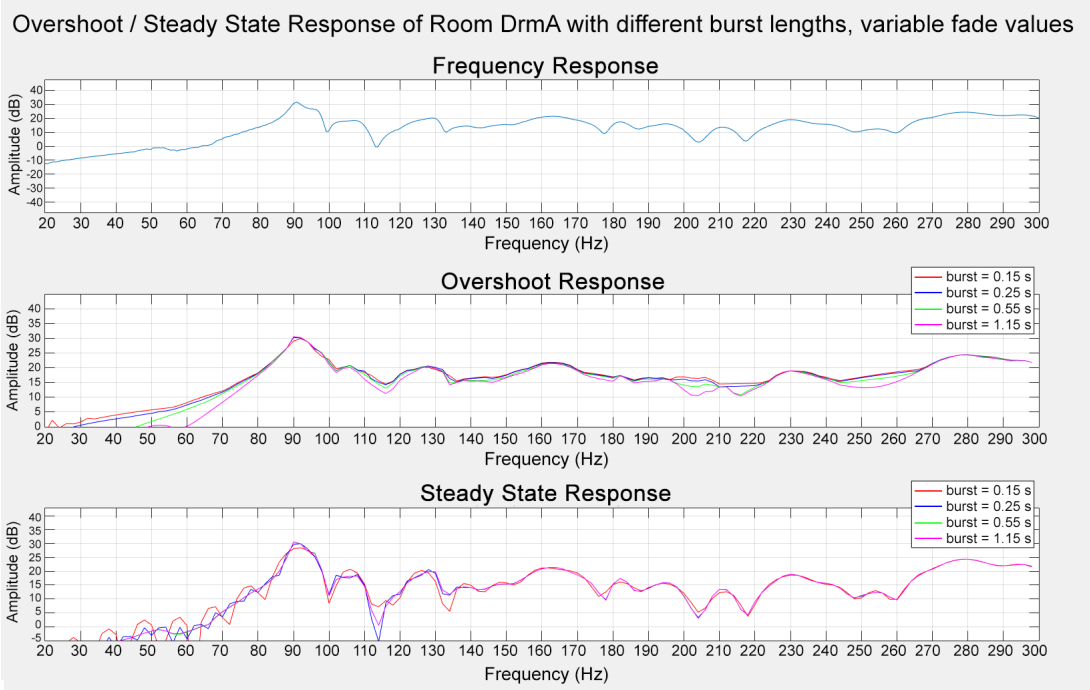
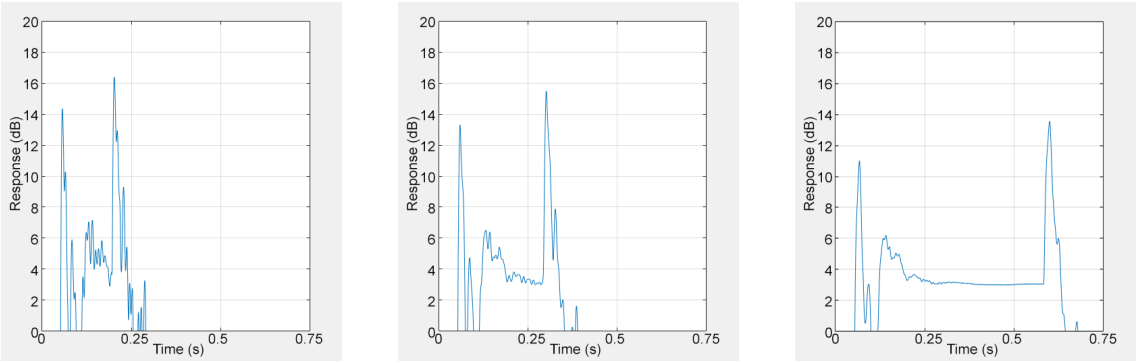


Figure 5.20: Room DrmA response where fade-in and fade-out lenghts are 3 percent of the burst length



(a) 150 ms burst (b) 250 ms burst (c) 550 ms burst

Figure 5.21: Different burst lenght with fade in/out values = 3 percent of the burst length at 204 Hz in room DrmA, resulting in different overshoot amplitude

Intuitively, if the attack time of the note changes, the overshoot will change as well. The quicker the attack, the higher the overshoot.

It is clear that, if the variable under test is the note length, the attack time does not differ, (e.g., a short bass note and a long bass note should have the same attack) therefore the overshoots should have the same amplitude. Fade-in and fade-out values of constant duration have indeed solved this problem.

Constant fade duration

Initially, the number of samples corresponding to the 3 percent of 0.15 seconds was used as the fixed value to test if the overshoot would have the same amplitude. Such value results in 0.0045 seconds, which corresponds to 199 samples at 44.100 Hz, In fact:

$$(0.15 * 44100)/100 * 3 = 198.45$$

At this stage, the duration of the attack and decay part of real world sounds (kick notes and bass notes) was not taken into account yet. This analysis was done in order to see if the overshoot amplitude would be the same with constant fade-in and fade-out values over different burst durations.

Testing the same scenario with constant fade-in and fade-out values for all bursts' duration, it is clear that the overshoots now reach the same amplitude, as shown in the central subplot of Fig. 5.22 and in Fig. 5.23. The small difference in Fig. 5.22 between overshoot amplitude on peaks is due to the fact that the overshoot is defined as the maximum of the envelope and that really short signals do not always reach their steady state on frequency response's peaks.

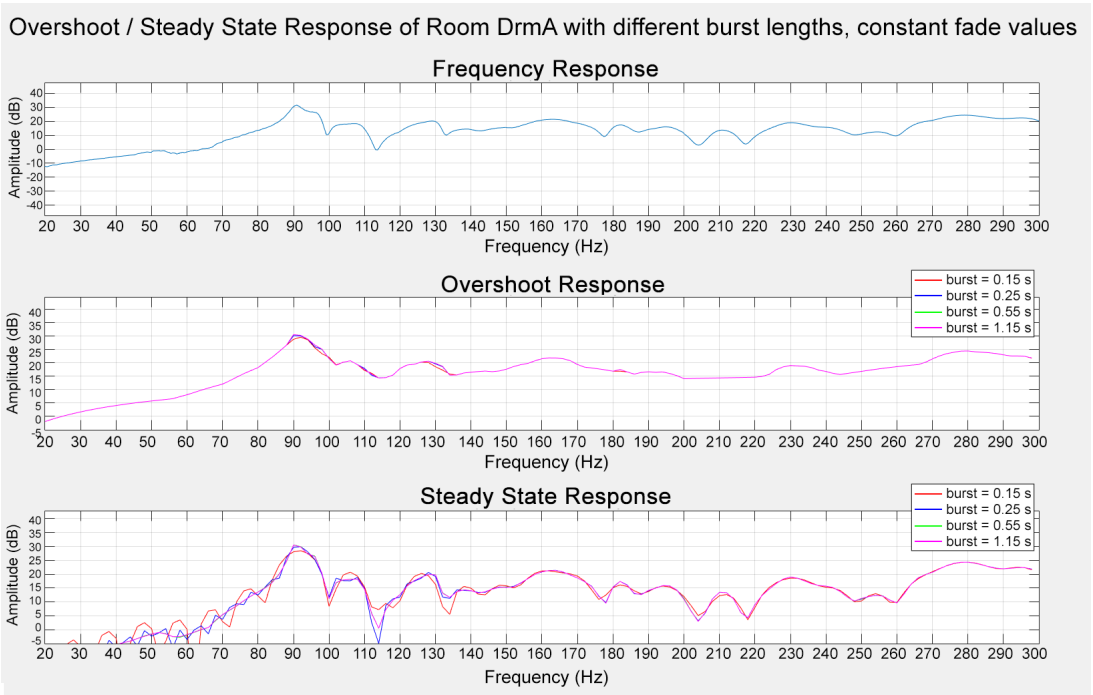
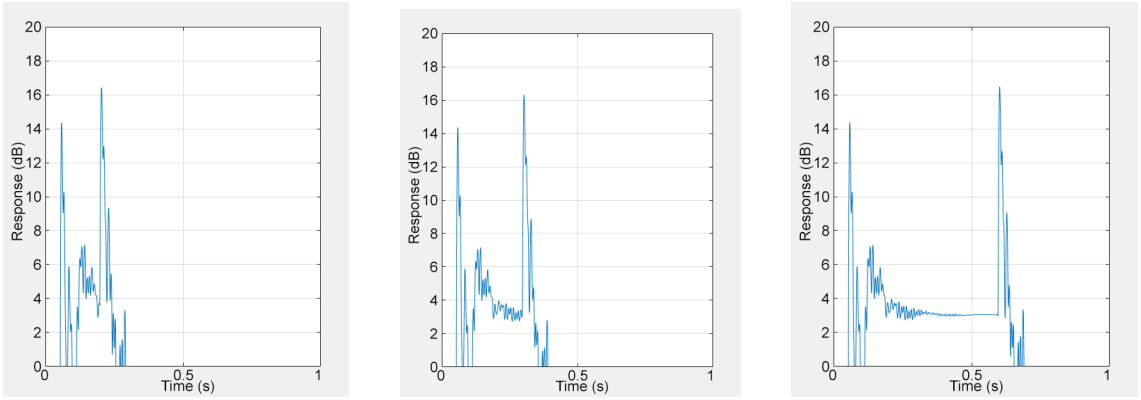


Figure 5.22: Room DrmA response with constant 199 samples fade-in/out



(a) 150 ms burst (b) 250 ms burst (c) 550 ms burst

Figure 5.23: Different burst lenght with constant 199 samples fade-in/out at 204 Hz in room DrmA, resulting in overshoots with the same amplitude

5.10 Room temporal domain simulation

The previous test made clear that the AQT algorithm should run with constant fade-in and fade-out values in this scenario.
The next step is deciding which values of fade-in and fade-out to use in the AQT simulation for simulating real world kick and bass sounds.

At this point, it is clear from the previous section that the overshoot behavior (i.e., the presence of overshoots and their amplitude) does not change with respect to the note duration. However, it changes with respect to the initial slope of the envelope of the triggering sound: if it is a step signal, the overshoot will be very high. On the other hand, if it is a slowly growing signal, the overshoot will be low. The type of attack of the note, therefore, the slope of the envelope, may influence the perception of the attack and volume (this phenomenon is related to the opening behavior in damped resonant systems of higher order). Therefore, to proceed with the analysis, it is important to arbitrarily choose such values. This could be an interesting area for further research.

There is no exact way to determine the correct value of fade-in and fade-out for real world kick and bass sounds, so a plausible value can be determined by averaging different versions of such sounds.

It is important to understand that this decision has been made only to facilitate the simulation. In fact, it does not impact the psychoacoustic tests, as those will be performed with kick and bass sounds (instead of test tones) on which it is not needed to set a fade-in and fade-out time, as they are already featured in their envelope.

The aims of this analysis were: running the AQT algorithm with plausible values in order to spot critical frequencies, to analyze overshoot behavior with envelopes similar to the real world ones, and to determine if the ending overshoot is present also on impulsive sounds. Therefore, linear slopes have been used for fade-ins and fade-outs for the sake of simplicity. For more advanced analyses, further studies could be conducted using more detailed envelopes in order to deliver even more realistic simulations directly in the AQT algorithm environment.

Sound	Attack time [s]
Kick 1	0.005
Kick 2	0.004
Kick 3	0.007

Table 5.1: Kick attack time

5.10.1 Kick sounds simulation

Kick sounds have a large peak at their beginning, but the sound is never sustained so the envelope of the test tones has to follow this behavior to correctly simulate kick notes. For this research, the author used synthetic kick drums from the virtual instrument "Sonic Academy Kick - Nicky Romero Edition™" that allows the user to tune the fundamental of the note. In order to understand the fade-in times of typical kick sounds, some kick sounds from well known sample libraries (Toontrack Superior Drummer™, Steven Slate Drums Platinum™, IK Multimedia SampleTank 2™) were played and rendered. These sounds were normalized and the attack time was chosen as the time between the first nonzero sample and the peak of the sound. The results are collected in table 5.1.

For the simulation, 0.005 seconds was chosen as the fade-in volume, since it is close to the medium value of the three attack times. As for the decay, a linear slope was used from the peak to the end of the note (which had length 150 msec for tests aimed at simulating kick sounds). The fade-out length is therefore equal to $150 - 5 = 145$ msec, which corresponds to 6394 samples at 44.1 kHz.

An exponential decay would have been closer to a real kick envelope, but this test aimed only at confirming that the influence of the closing overshoot is limited with non-sustained sounds.

Fig. 5.24 shows the envelope with the chosen values for the kick simulation over a 150 msec duration.

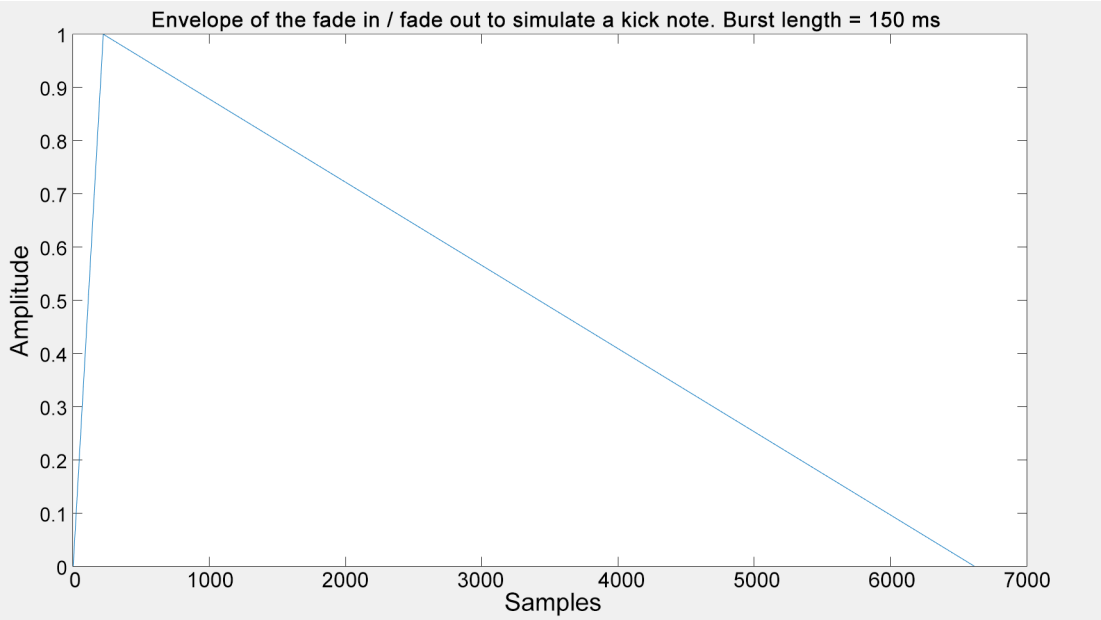
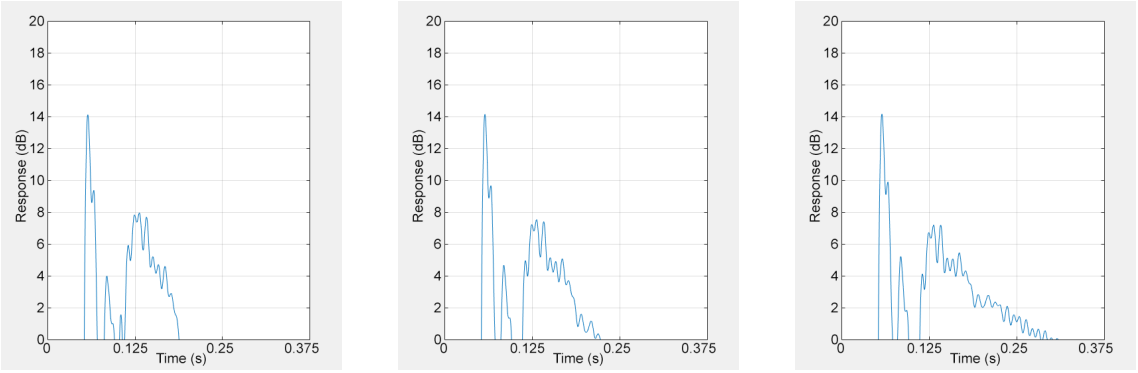


Figure 5.24: Sine wave envelope chosen for kick sound simulations

This analysis was carried out with the chosen burst lengths to highlight this behavior. However, the most meaningful one is the 150 msec test, since the sound is impulsive.



(a) 150 ms burst (b) 250 ms burst (c) 550 ms burst

Figure 5.25: Different burst lenght with kick envelope at 204 Hz in room DrmA

The AQT analysis with Kick envelope (shown in Fig. 5.25 with the same room and frequency of Fig. 5.23) shows that the opening overshoot behavior is the same with respect to the sustained sounds. However, with the Kick envelope, the closing overshoot behavior is not present. This is also intuitive because, since it is not a sustained sound, the volume starts to drop after the initial attack so when the direct sounds stop the reflected sound does not have an impulsive behavior, like it does with sustained envelopes.

Therefore, it is important to remember that, for non-sustained impulsive sounds, only the opening overshoot is meaningful.

5.10.2 Bass sounds simulation

Since bass notes are sustained sounds, the envelope is quite similar to the envelope of pure tones. However, bass notes feature an attack time and a release time which have to be taken into account when shaping the test tones' envelope. For this research, the author used a bass library from a well-known sound library. A 0.15 seconds note of Fretless bass, Fingerstyle bass, Picked bass and Slap bass were played and rendered at the same volume. These sounds were normalized and the attack and decay time was chosen as follows:

The author chose the attack time as the time between the first nonzero sample and the peak of the first part of the envelope.
The decay was chosen as the time ranging from the instant where the note stopped to the moment the envelope arrived near zero. Small fluctuations at the end were ignored.
This analysis just aimed at giving a rough idea of the attack and decay time for this type of instruments played in different ways. The results are collected in table 5.2.

For the simulation, 0.005 seconds (corresponding to 221 samples at 44.1 kHz) was taken as the fade-in value, since it is close to the median value of the four attack times, and since it is the same as the chosen fade-in value for kick sounds. This allows for a faster preliminary analysis. The chosen decay time for simulating

Sound	Attack time [s]	Decay time [s]
Fretless Bass	0.002	0.017
Fingerstyle Bass	0.012	0.023
Slap Bass	0.013	0.020
Pick Bass	0.001	0.017

Table 5.2: Bass attack and decay time

bass sound envelopes was 0.020 seconds (corresponding to 882 samples at 44.1 kHz), which is close to the median value of the decay times. Fig. 5.26 shows the envelope with the chosen values for the bass simulation over a 150 msec duration.

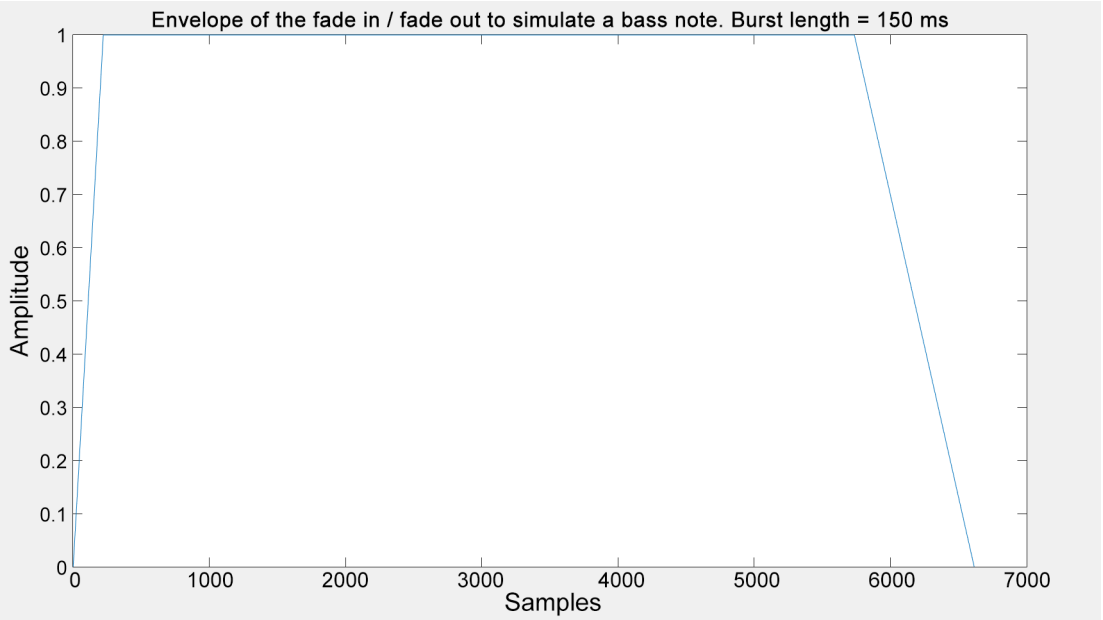


Figure 5.26: Sine wave envelope chosen for bass sound simulations

The AQT simulation with bass envelopes and burst length 150, 250 and 550 msec is what is used in the following steps of this research (thanks to the fact that the same fade-in length was chosen for both kick and bass simulation). It is important to always remember that the closing overshoot is significative only for sustained sounds, since the lack of energy in the last part of the decay of an impulsive sound does not generate an overshoot at the end of the AQT Response Envelope, as explained earlier, and as shown by Fig. 5.25.

5.10.3 A word of caution regarding Room Slowness with non-sustained sounds

The room Slowness parameter, as already stated, is closely correlated to the room Inertia values. However, the temporal behavior analysis was performed, as already stated, for sustained short notes. It is important to understand that, when the note is not sustained (such as, in kick drum hits), Room Slowness fails to completely describe the speed of the AQT Response Envelope in reaching its steady state value, since, by definition, a steady state value is not present in impulsive sounds. When the AQT Response Envelope of an impulsive sound starts to grow, frequencies with

higher Slowness are indeed slower, but the difference is minimal if compared to the same behavior in a sustained sound. The sustained sound behavior is shown in Fig. 5.27, where the AQT Response Envelope of two different frequencies is shown (Blue line corresponds to 32 Hz after convolution, red line corresponds to 38 Hz after convolution, while the other lines are the envelope of both notes before convolution) in room PRZ. The different speed in reaching the steady state value is evident.

Fig. 5.28 instead shows the same behavior when tested with the Kick envelope. It is clear that the behavior is the same, but the fact that the sound is not sustained minimizes the impact of the Slowness variable. As the figures make clear, Slowness remains an interesting parameter for sustained sounds.

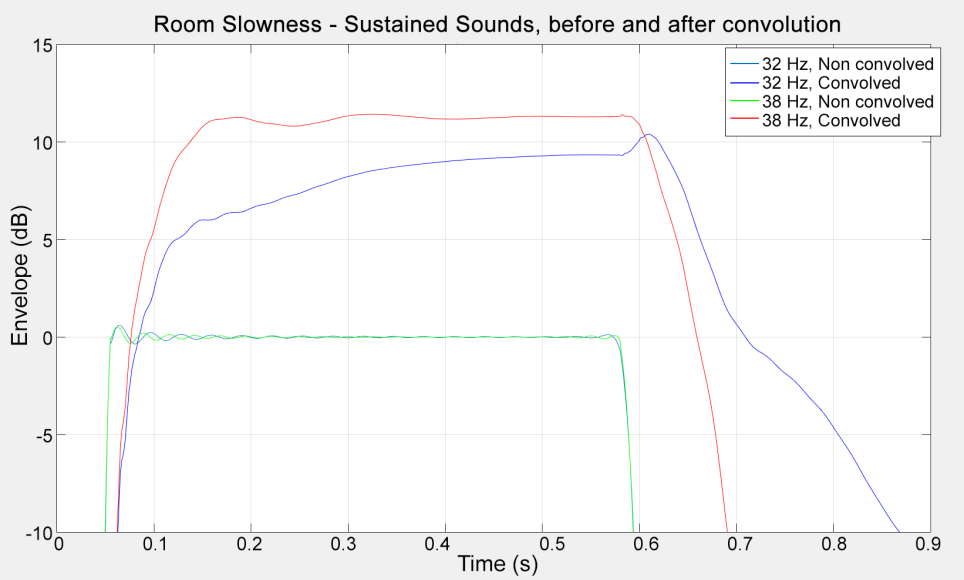


Figure 5.27: Room Slowness effect on a sustained sound

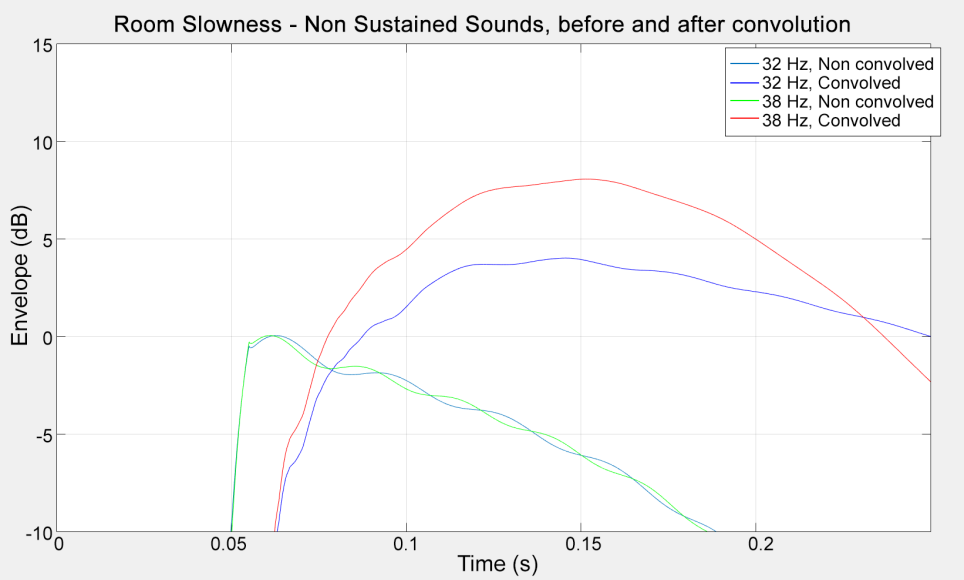


Figure 5.28: Room Slowness effect on a non-sustained sound

5.11 Room AQT Analysis

The AQT analysis was performed for each of the 8 real rooms under test. Each room was excited with bursts of the chosen lengths (150, 250, 550 milliseconds) with the chosen fade-in and fade-out values (”bass” envelope). The frequency response and EFT plot were checked to find problematic frequencies (peaks and valleys in the frequency response) and the AQT Response Envelope of those frequencies were checked. The results are as follows, for each room. The reader is reminded that Temporal Behavior and Advanced Overshoot Analysis blocks were added after the first psychoacoustic tests and before the second psychoacoustic test.

5.11.1 Room CN

Room CN is a simmetric mixing room with angled roof and dimensions 3.7, 2.9 and 2.9 meters, with a volume of 30.6 m^3 and plastered masonry, highly reflective surfaces.

Figs 5.29, 5.30, 5.31 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room CN. Frequency response peaks are present at 44, 76, 86 Hz, which do not reach steady state with 0.15 s bursts, and at 192, 230 Hz, which instead do. Valleys at 60, 82, 96 Hz show moderate opening overshoot and little closing overshoot.

While this room does not feature strong Overshoot behavior, it reacts very slowly to frequencies on the frequency response’s peaks, returning high values for the Room Slowness parameter. Such frequencies (as an example, Fig. 5.2) are also very slow when decaying. This is one of the rooms with higher values for Room Slowness, Decay Time and Room Inertia. Therefore, this will be used as a ”slow” room when comparing effects of different Slowness and Inertia values.

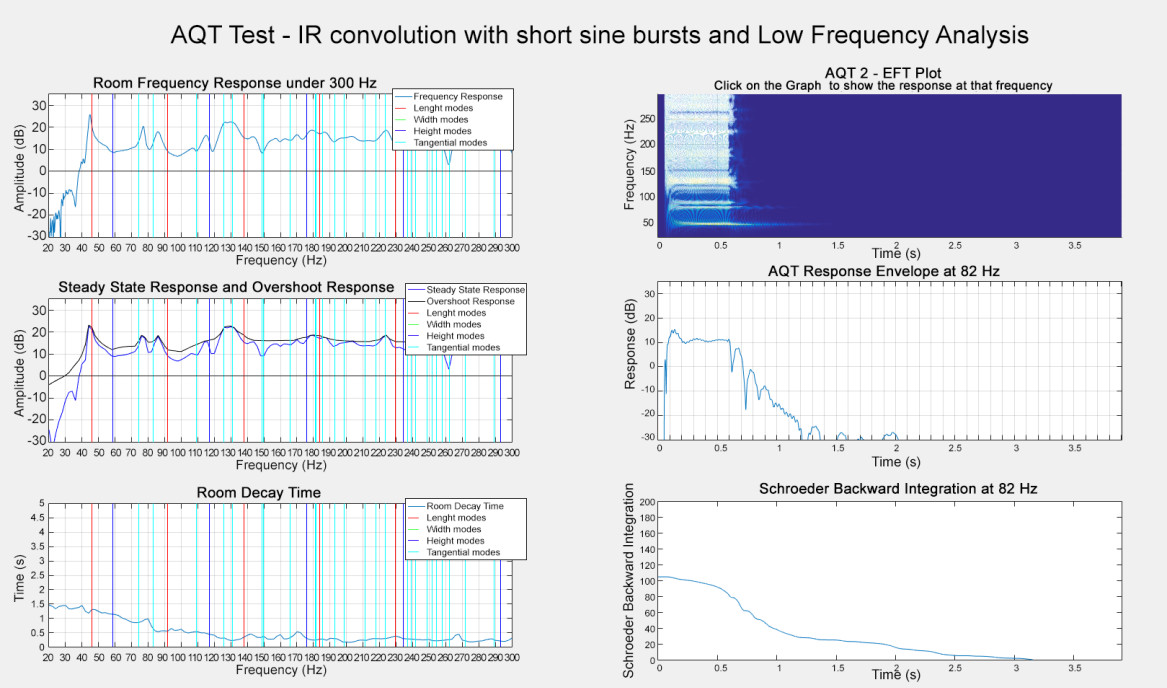


Figure 5.29: Room CN - AQT Results Plot

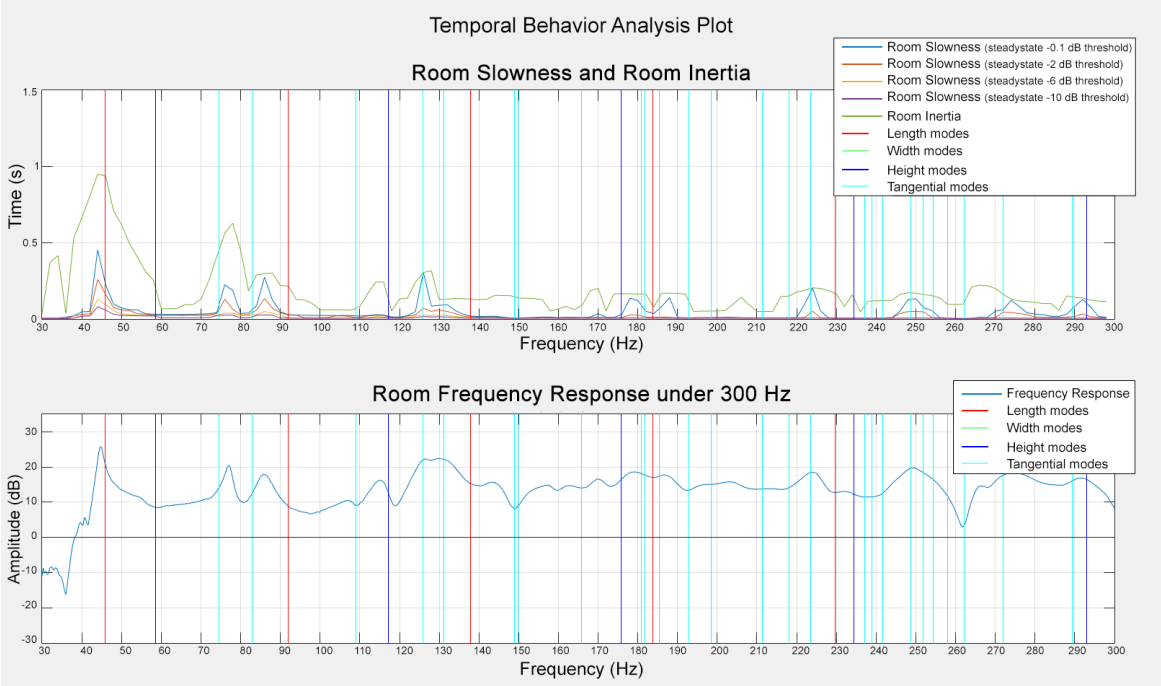


Figure 5.30: Room CN - Temporal Behavior Analysis Plot

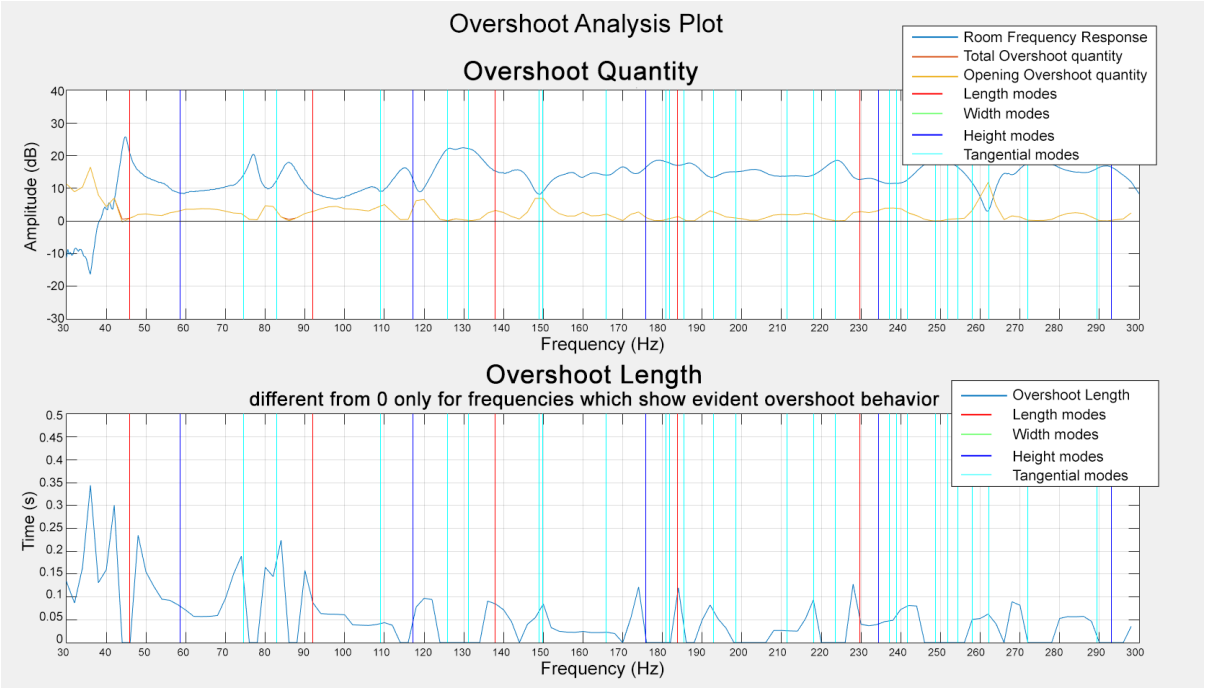


Figure 5.31: Room CN - Overshoot Analysis Plot

5.11.2 Room PRZ

Room PRZ is an irregular but simmetric mixing room with dimensions 5, 3 and 2.5 meters, with a volume of 38.4 cubic meters, gypsum boards surfaces treated with sound absorbing mats.

Figs 5.32, 5.33, 5.34 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room PRZ.

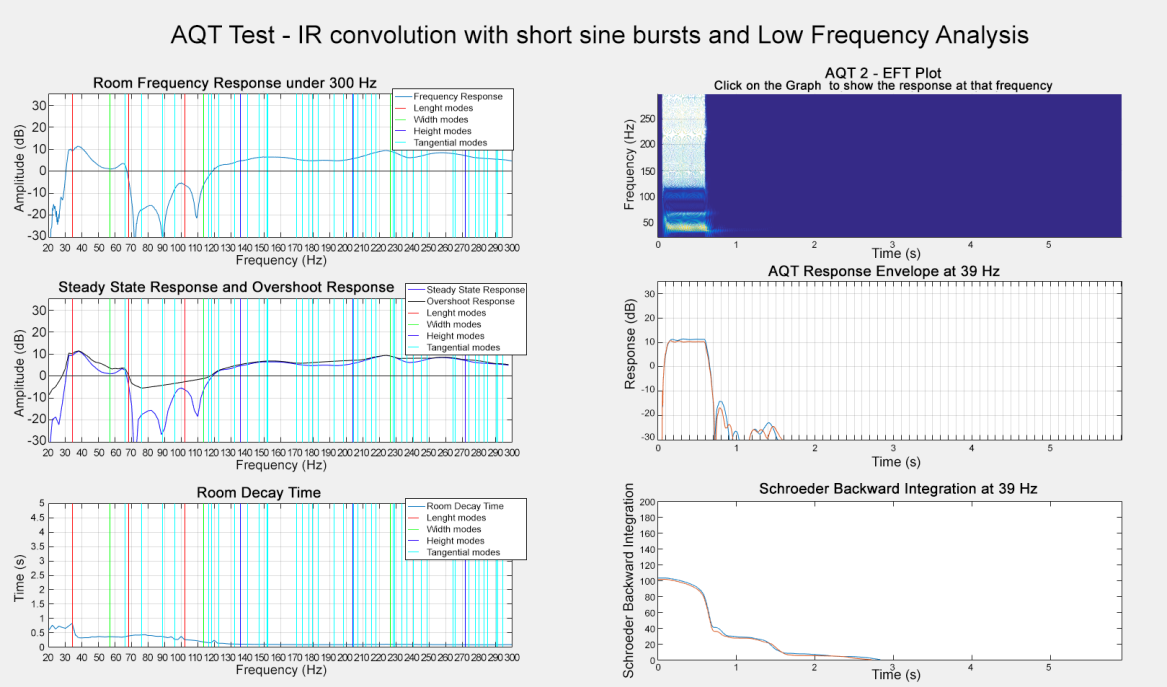


Figure 5.32: Room PRZ - AQT Results Plot

Frequency response peaks are present at 38, 64, 154, 224, 258 Hz (reaching steady state values even with 0.15 s bursts). The most significative valleys have their minimum value at 58 Hz showing low overshoot behavior, 72, 110 Hz (Fig.5.4) showing very strong overshoot behavior.

This is one of the most interesting rooms in this research. As the AQT plots show, this room features a huge valley in its frequency response, indicating that it lacks in low frequencies between 70 and 120 Hz. However, the overshoot behavior in the same frequency range is very high (Fig.5.34, Fig. 5.4), suggesting the hypothesis that, for short sounds, the listener could grasp the overshoot behavior instead of the steady state one, making less evident the perception of the huge frequency response valley of this room. Another very important feature of this room is that its frequencies, also on low frequency room modes, reach their steady state value very fastly (see Fig. 5.3). They also decay very fastly, resulting in very low values for Room Slowness, Room Decay Time and Room Inertia. Another interesting feature in this room is that the peaks at 32 and 38 Hz feature very different Room Slowness and Inertia values, 32 Hz having the higher ones despite having a minor amplitude than 38 hz.

These facts make this room a great candidate to test the effects of both Overshoot Behavior and low values of Room Slowness and Room Inertia. Because of the huge frequency response valley, it is also a good room for testing frequency response perception, as introduced by [37].

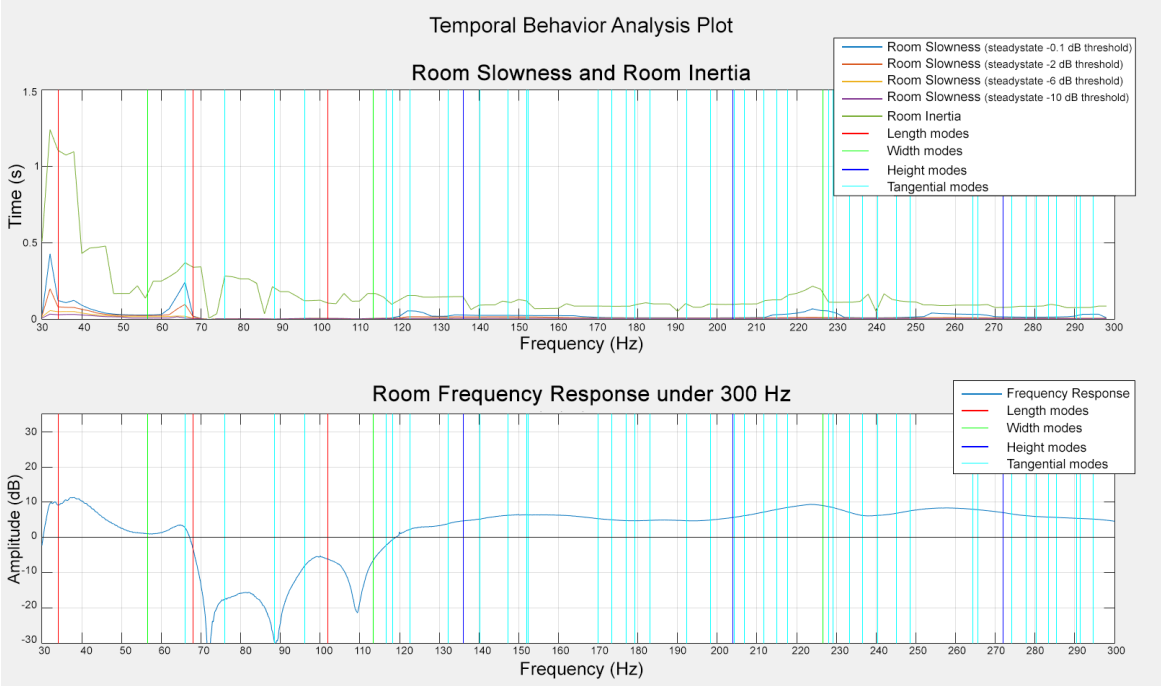


Figure 5.33: Room PRZ - Temporal Behavior Analysis Plot

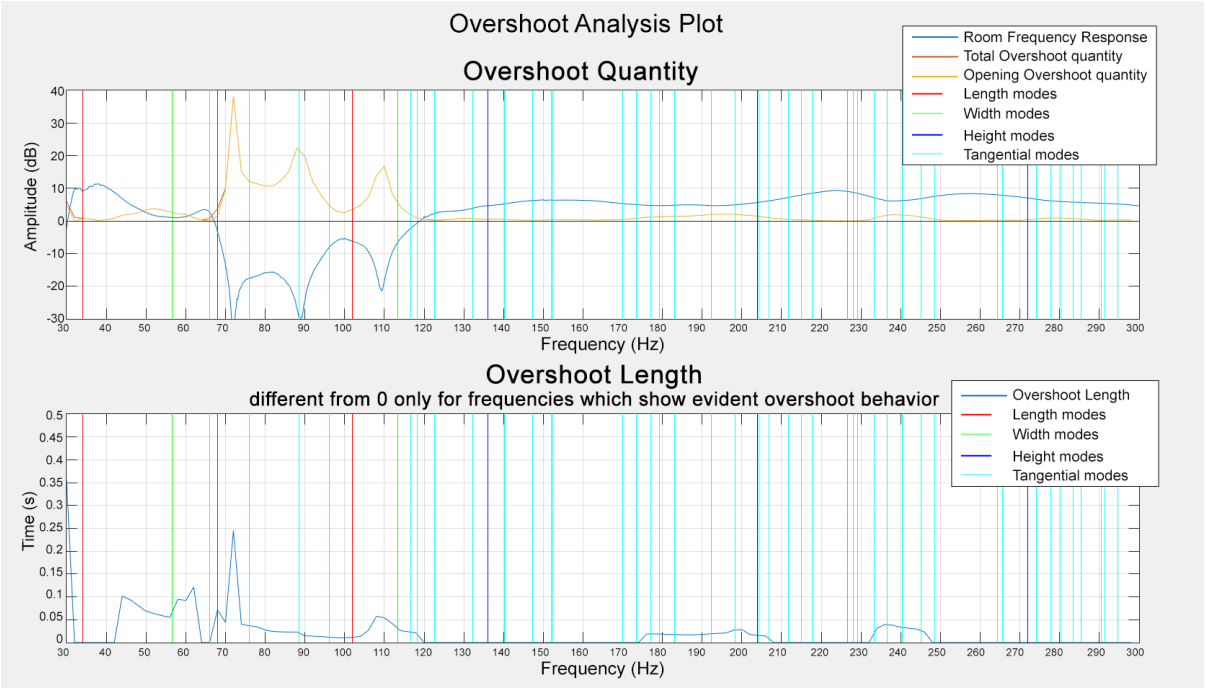


Figure 5.34: Room PRZ - Overshoot Analysis Plot

5.11.3 Room SGR

Room SGR has dimensions 4.6, 3.6 and 2.6 meters, with a volume of 43.1 cubic meters. It is a parallalepiped simmetric room and it is used as a mastering room. The walls are gypsum boards with specific correction.

Figs 5.35, 5.36, 5.37 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room SGR.

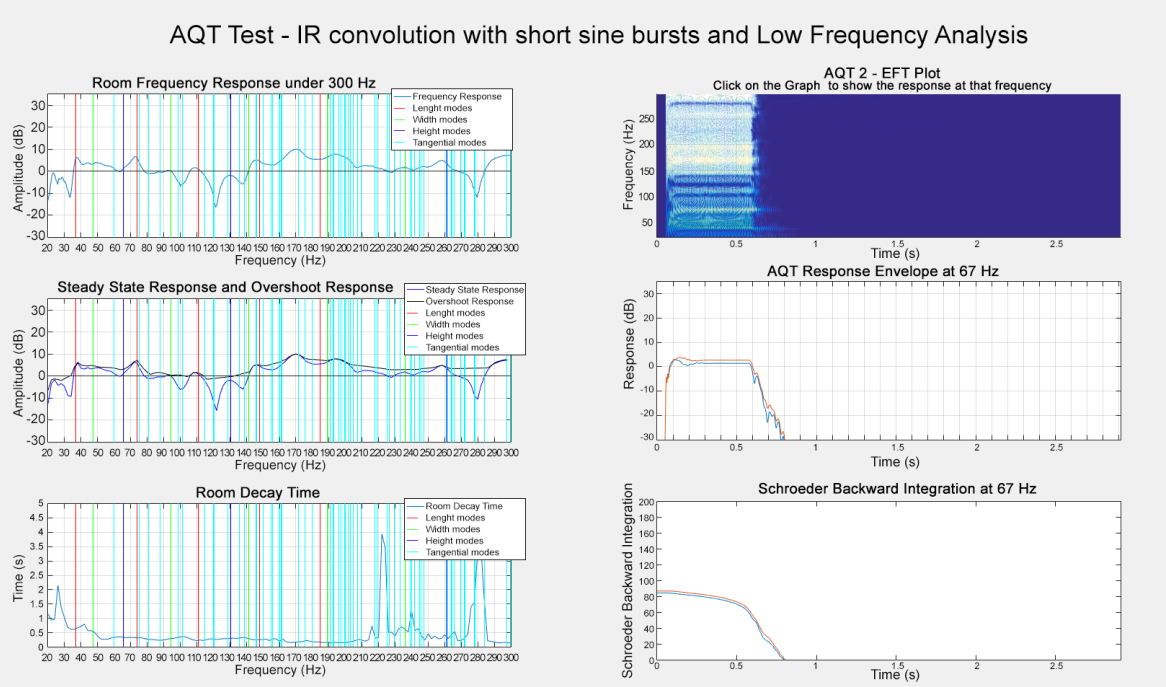


Figure 5.35: Room SGR - AQT Results Plot

Room SGR has a peak at 38 Hz which does not reach steady state with a 0.15 s tone, peaks at 108 and 170 Hz which instead do, valleys at 64 and 100 Hz with strong opening overshoot and small closing overshoot, valleys at 122 and 280 with strong overshoot behavior.

As far as temporal behavior goes, this room has a slightly higher Slowness value than PRZ, but lower Inertia. This room will be very important in the second test, because it will be used to test those two parameters against room PRZ, and to test the perception of room PRZ's hole in the frequency response.

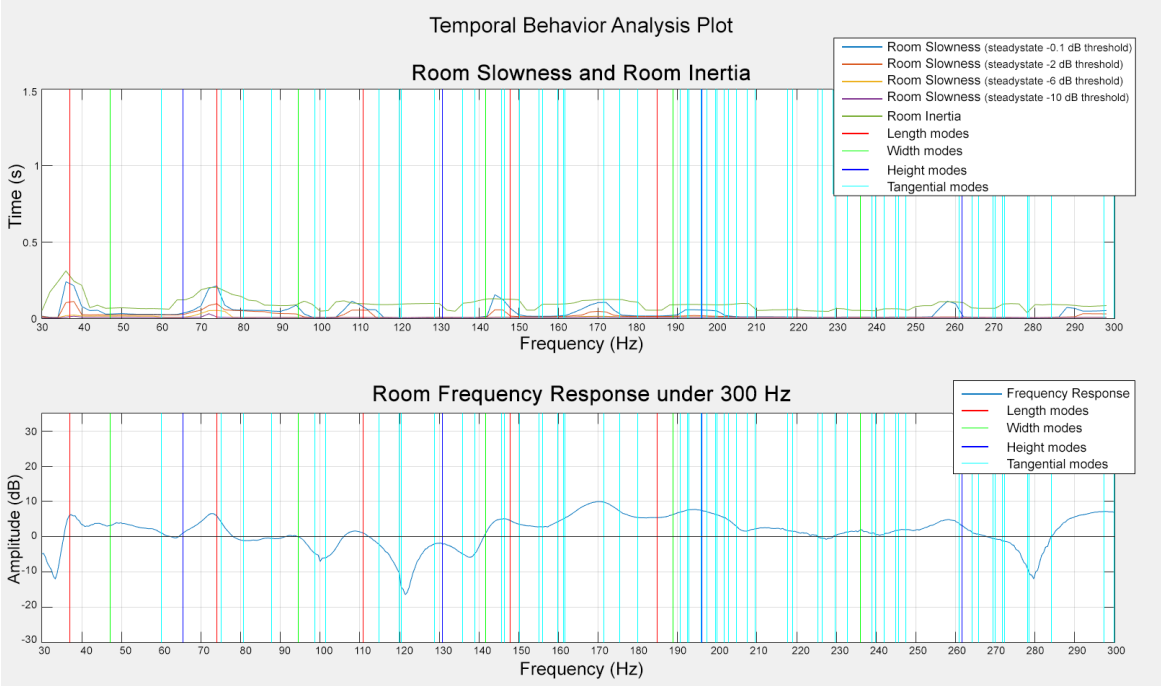


Figure 5.36: Room SGR - Temporal Behavior Analysis Plot

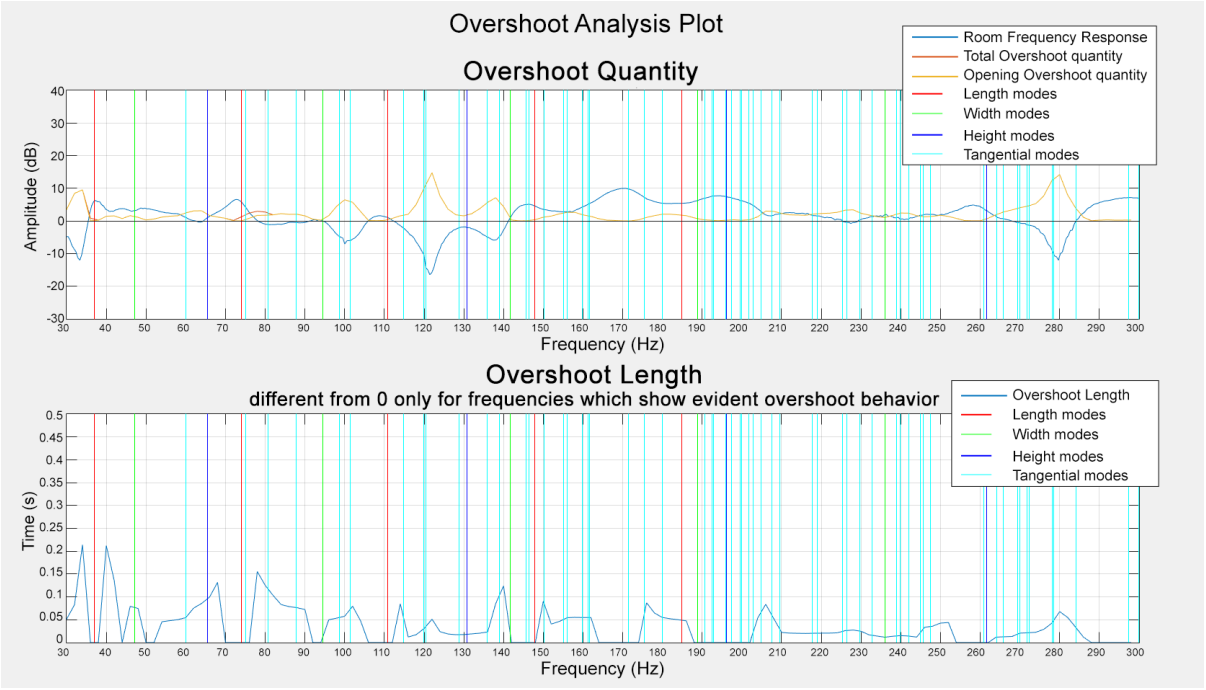


Figure 5.37: Room SGR - Overshoot Analysis Plot

5.11.4 Room GD

Room GD has dimensions 4.8, 4 and 2.9 meters, with a volume of 55.2 cubic meters. It is a parallelepiped simmetric room used as a mixing room. Its walls are gypsum boards with no acoustic correction.

Figs 5.38, 5.39, 5.40 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room GD.

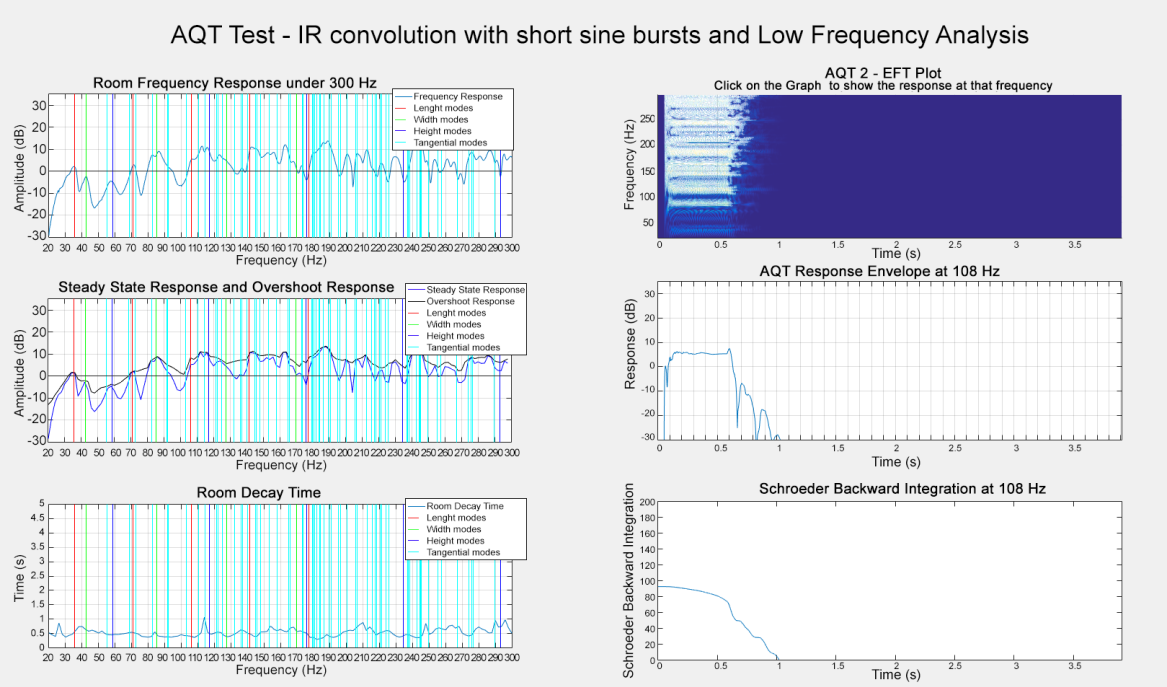


Figure 5.38: Room GD - AQT Results Plot

Room GD has an highly irregular frequency response behavior. Main frequencies of interest are peaks at 86 and 164 Hz, which do not reach their steady state value for 0.15 s test tones, and valleys at 134 Hz with strong opening overshoot, and at 176 Hz with strong closing overshoot.

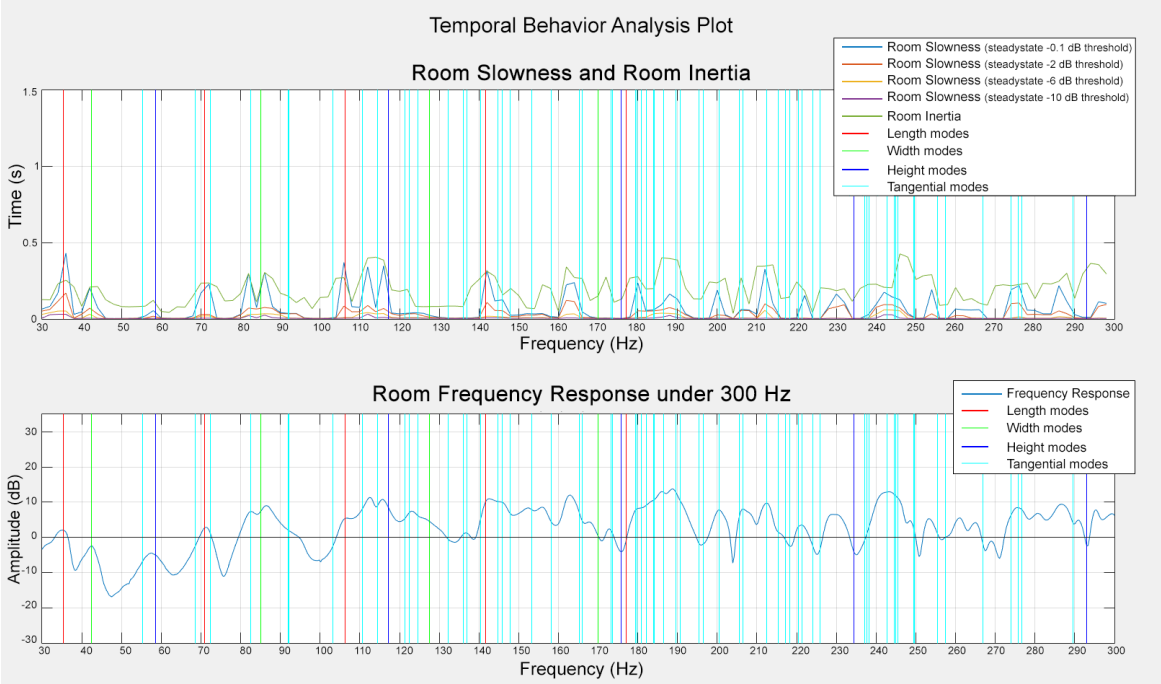


Figure 5.39: Room GD - Temporal Behavior Analysis Plot

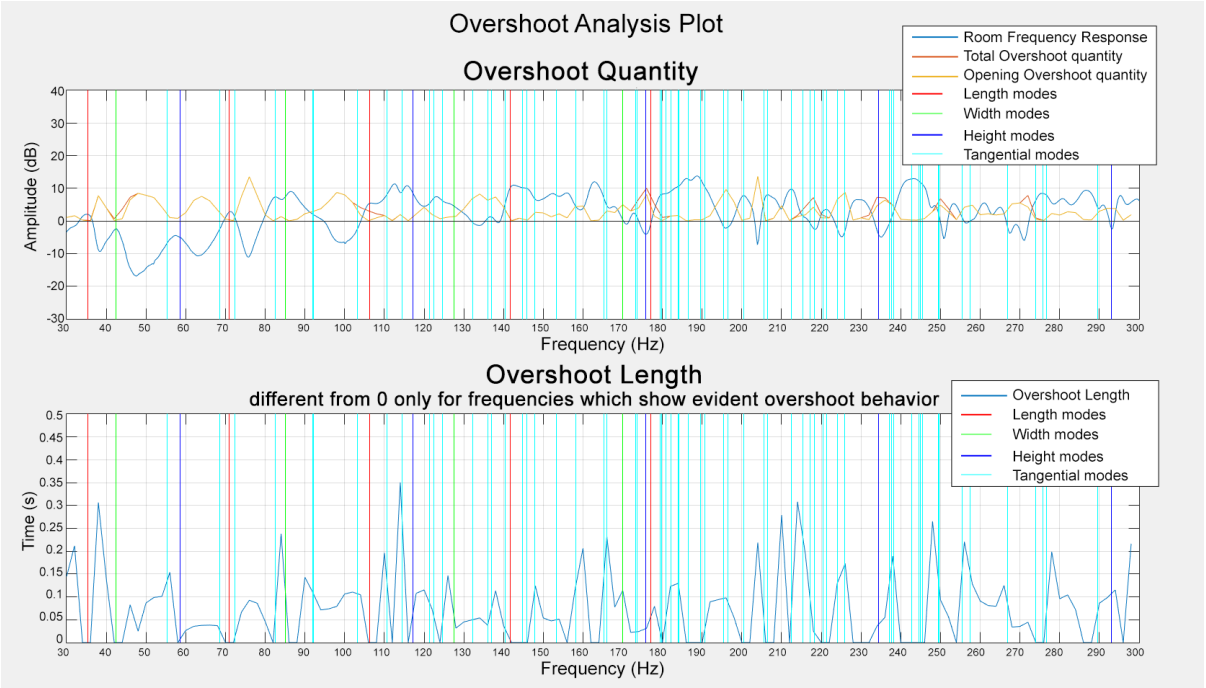


Figure 5.40: Room GD - Overshoot Analysis Plot

5.11.5 Room SNT

Room SNT has dimensions 7.2, 3.3 and 2.0 meters, with a volume of 46.7 cubic meters. It is a non simmetric room with angled roof used for hi-fi listening. Its walls are gypsum boards with no acoustic correction.

Figs 5.41, 5.42, 5.43 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room SNT.

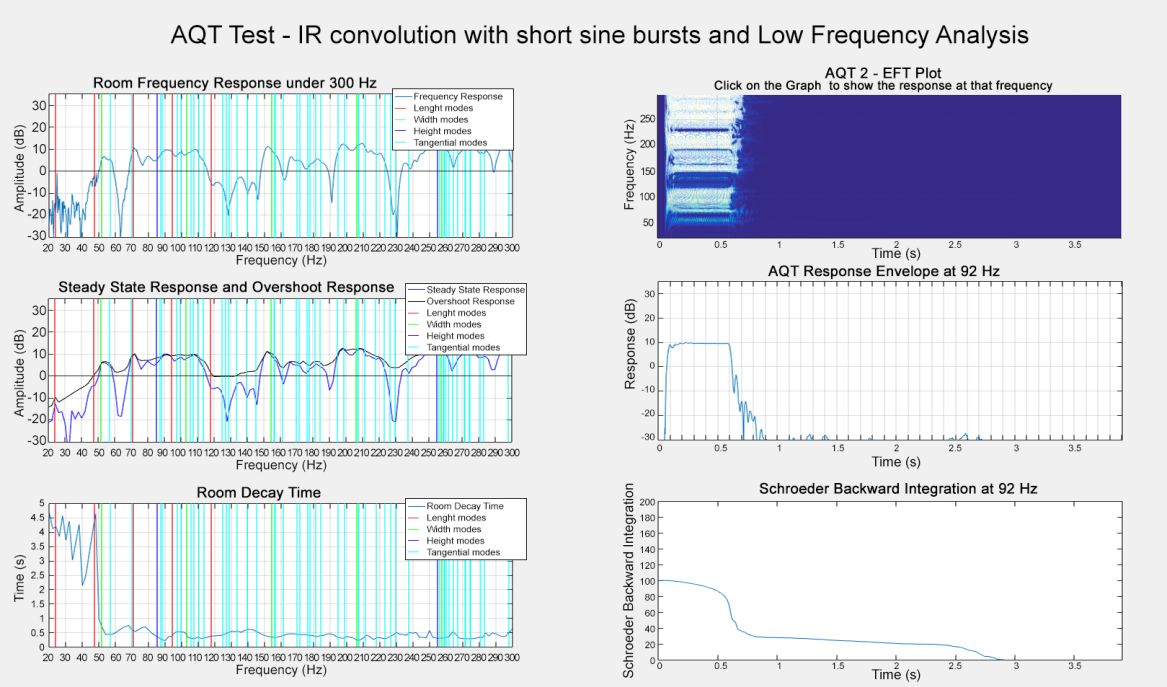


Figure 5.41: Room SNT - AQT Results Plot

Room SNT is very interesting because of its shape. It has a peak at 72 Hz that does not reach its steady state value for a 0.15 s test tone, peaks at 154 and 208 hz which instead do, valleys at 64, 128, 190, 230 hz with strong overshoot behavior and a valley at 162 with moderate opening overshoot and high closing overshoot.

Room SNT is the one that scored the highest values for Room Slowness, Decay Time and Inertia. It will be therefore used in the second test as a "very slow" room, which is supposed to impact very negatively the perceived precision and definition of sounds.

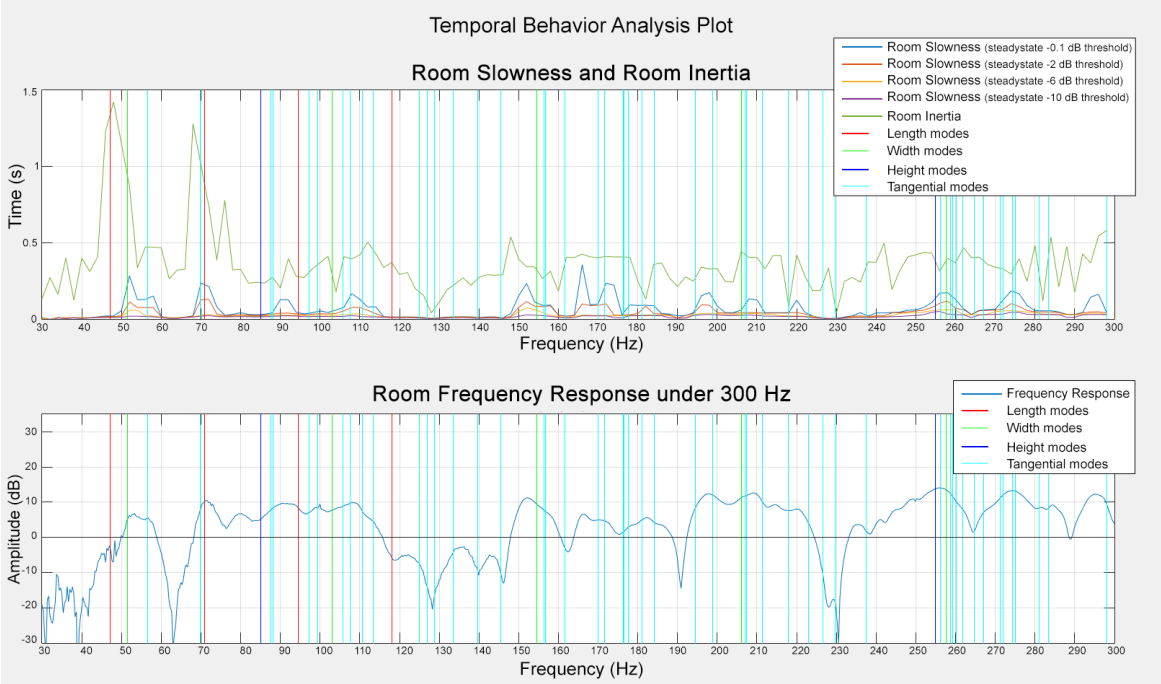


Figure 5.42: Room SNT - Temporal Behavior Analysis Plot

Figure 5.43: Room SNT - Overshoot Analysis Plot

5.11.6 Room DrmA

Room DrmA has dimensions 3.8, 3.5 and 2.8 meters, with a volume of 35.7 cubic meters. It is an irregular, non simmetric room used for recording speech and voice overs. Its walls are gypsum boards and the room is partially treated in the high frequency range.

Figs 5.44, 5.45, 5.46 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room DrmA.

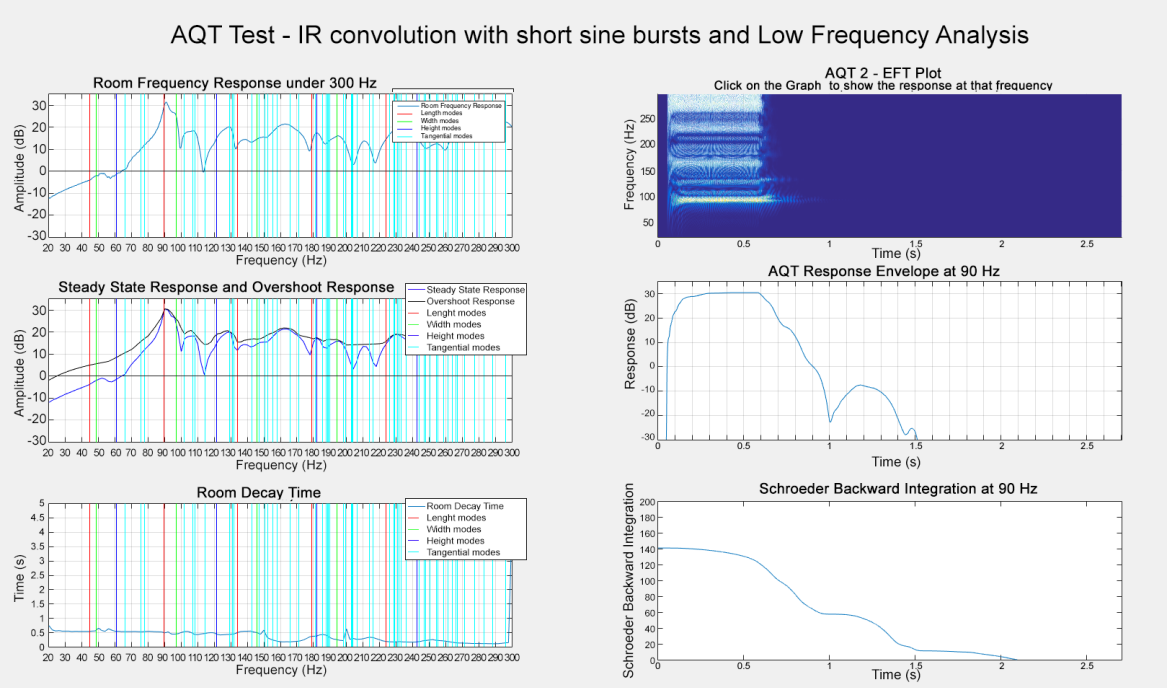


Figure 5.44: Room DrmA - AQT Results Plot

Room DrmA has a peak at 90 Hz that does not reach its steady state value for a 0.15 s tone, peaks at 164 and 280 Hz that, instead, do. It has valleys at 100 and 114 Hz with strong overshoot behavior and a valley at 218 Hz with strong opening overshoot and small closing overshoot.

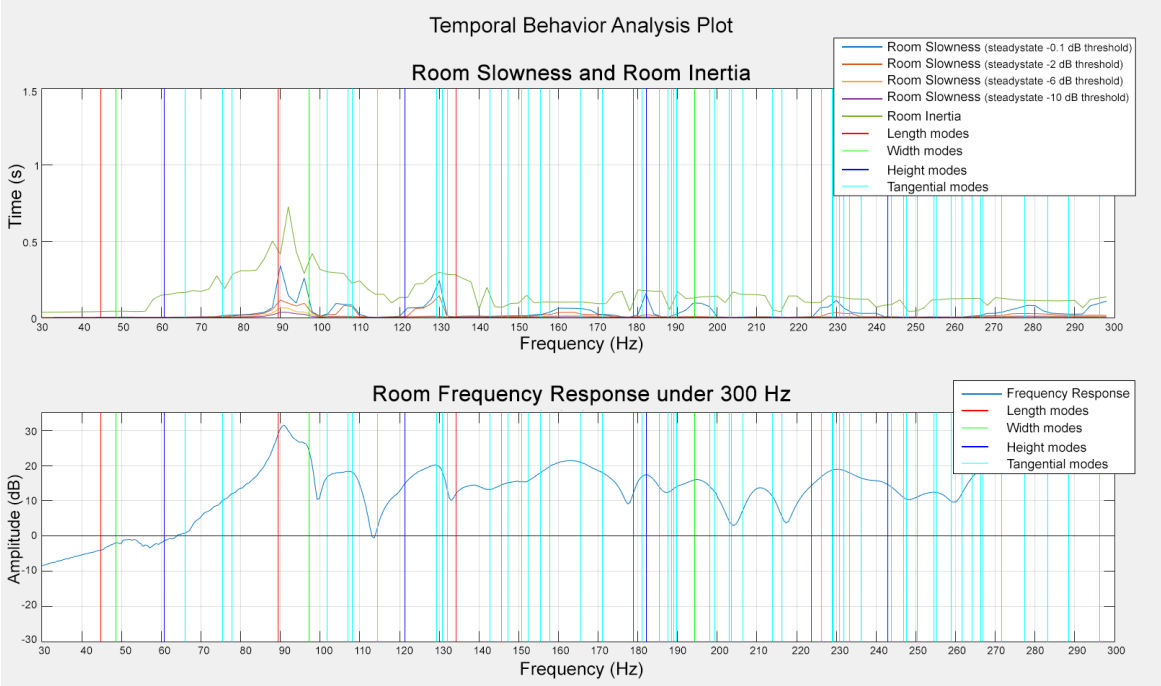


Figure 5.45: Room DrmA - Temporal Behavior Analysis Plot

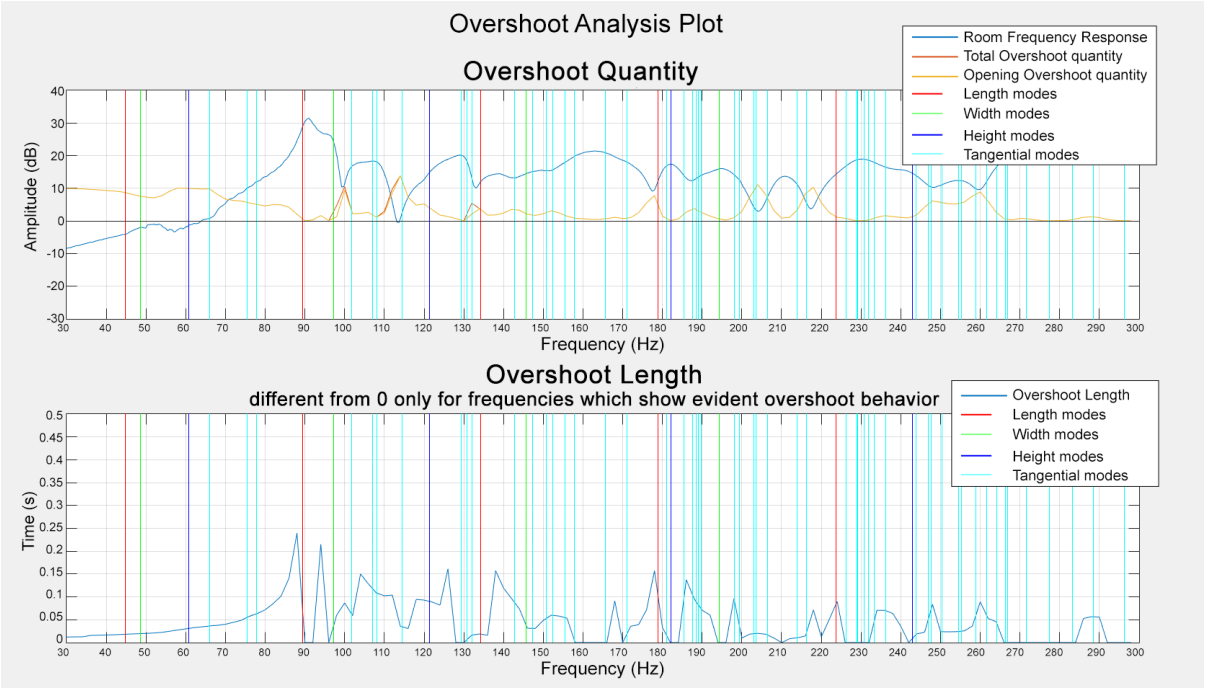


Figure 5.46: Room DrmA - Overshoot Analysis Plot

5.11.7 Room DrmB

Room DrmB has dimensions 3.6, 3.5 and 2.8 meters, with a volume of 34.6 cubic meters. It is an irregular, non simmetric room used for recording speech and voice overs. Its walls are gypsum boards and the room is partially treated in the high frequency range.

Figs 5.47, 5.48, 5.49 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room DrmB.

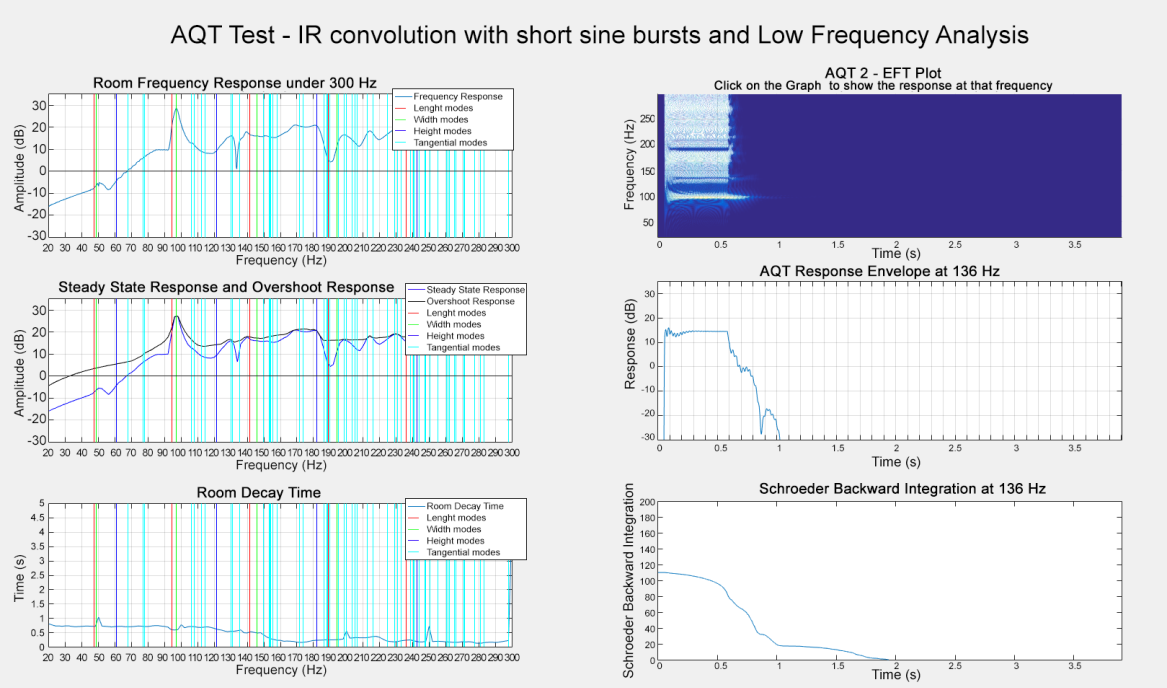


Figure 5.47: Room DrmB - AQT Results Plot

DrmB has peaks at 98 and 130 Hz, that do not reach their steady state value with a 0.15 s tone, peaks at 170 and 270 Hz that, instead, do. It also has a valley at 118 Hz with strong opening overshoot and small closing overshoot, and valleys at 134 and 190 Hz with strong overshoot behavior. It is possible to see how the square shape of the room has a visible effect on the frequency response: since two dimensions have very similar room modes, there will be high peaks in the frequency response as these modes originating from different room dimensions build on each other.

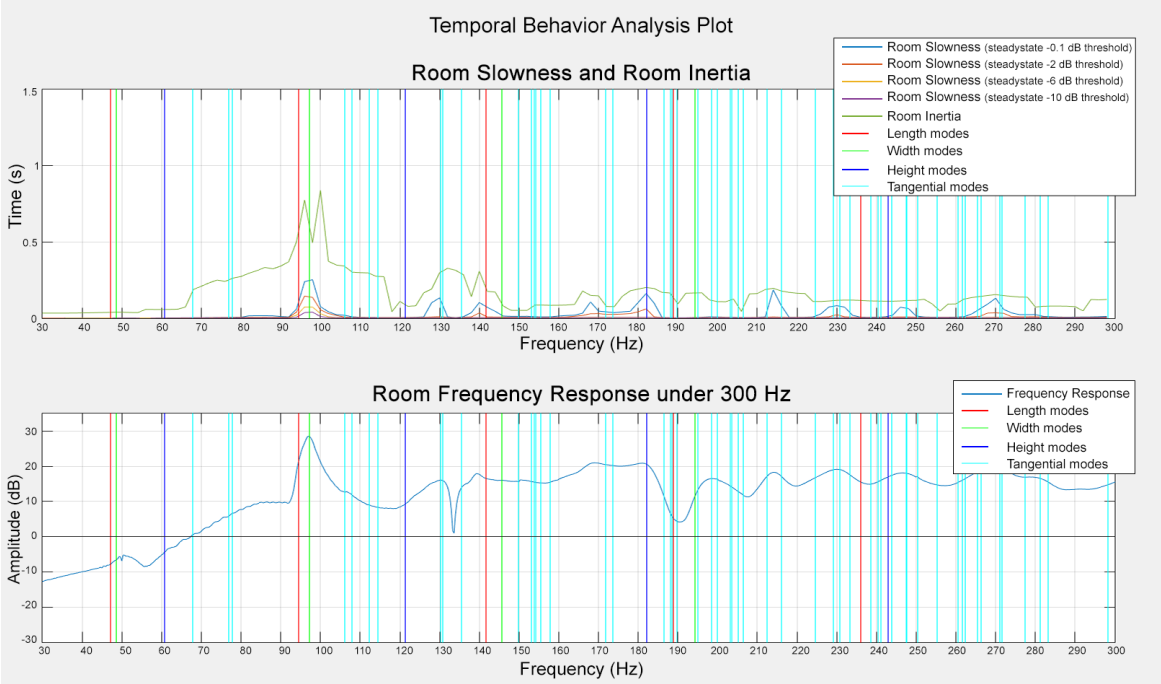


Figure 5.48: Room DrmB - Temporal Behavior Analysis Plot

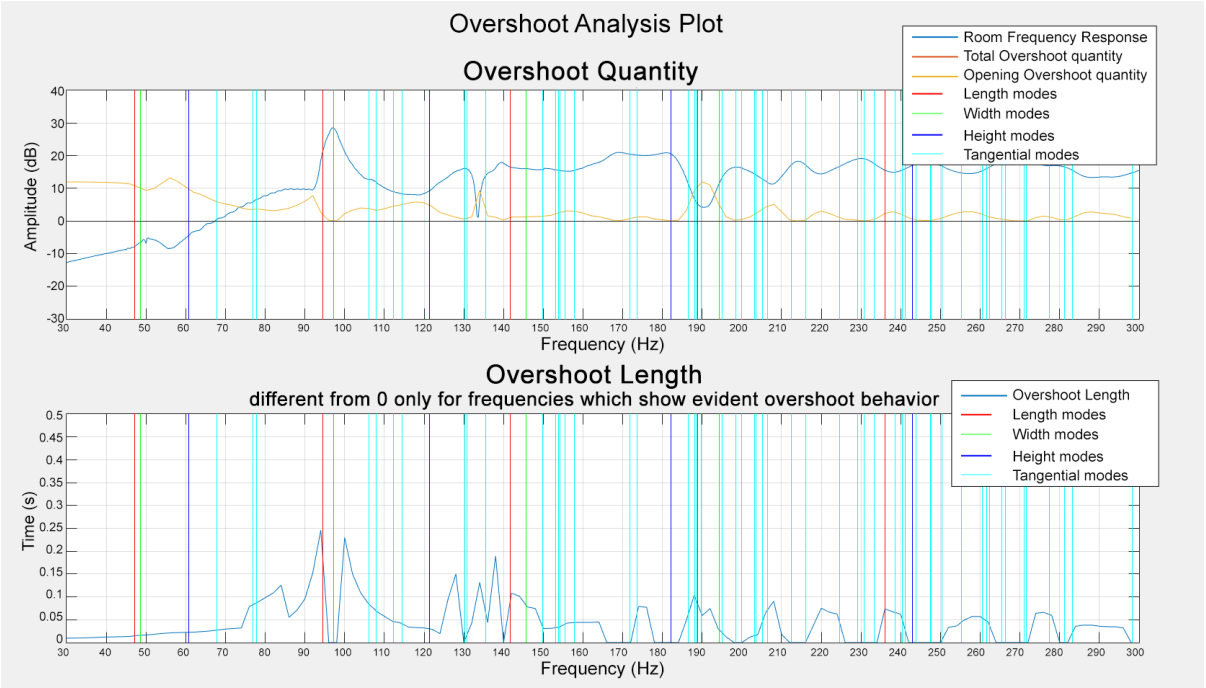


Figure 5.49: Room DrmB - Overshoot Analysis Plot

5.11.8 Room DrmReg

Room DrmReg has dimensions 4,0, 3.6 and 2.8 meters, with a volume of 39.6 cubic meters. It is a simmetric, parallelepiped room used for mixing speech and voice overs. Its walls are gypsum boards and the room is partially treated in the high frequency range.

Figs 5.50, 5.51, 5.52 show the AQT Results plot, the Temporal Behavior Analysis plot and the Overshoot Analysis plot for room DrmReg.

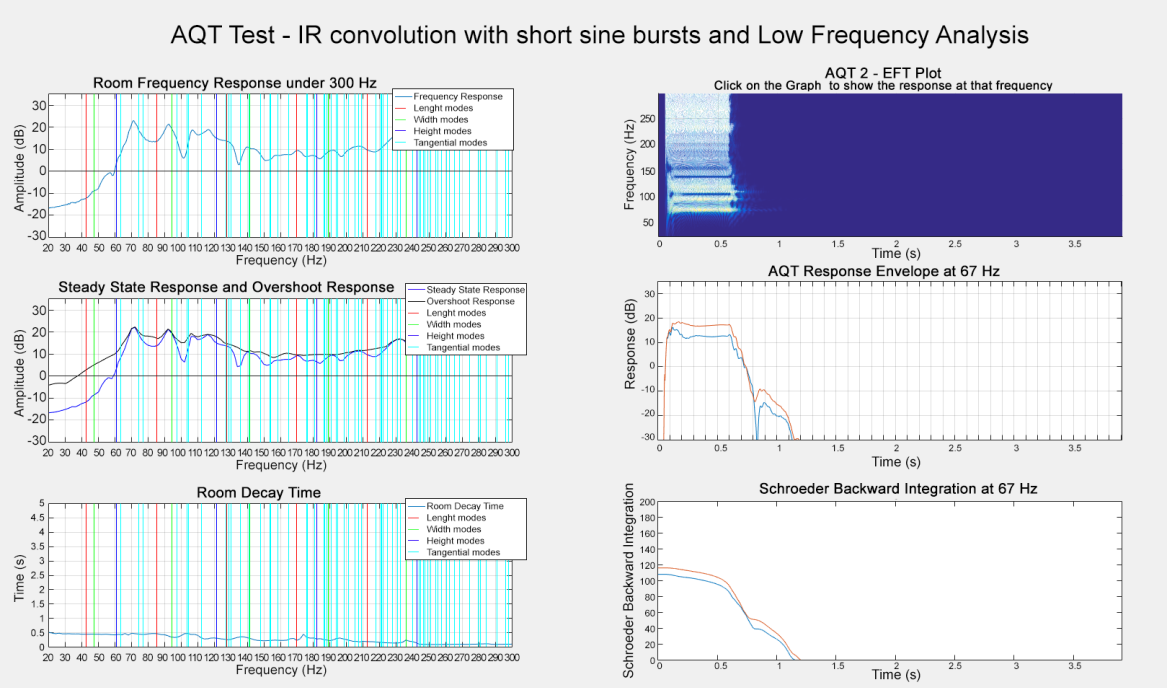


Figure 5.50: Room DrmReg - AQT Results Plot

DrmReg features peaks at 72 and 92 Hz that do not reach their steady state values with 0.15 s tones, peaks at 192 and 230 Hz that instead do, a valley at 102 with strong overshoot behavior, valleys at 134 and 286Hz with strong opening overshoot and small final overshoot.

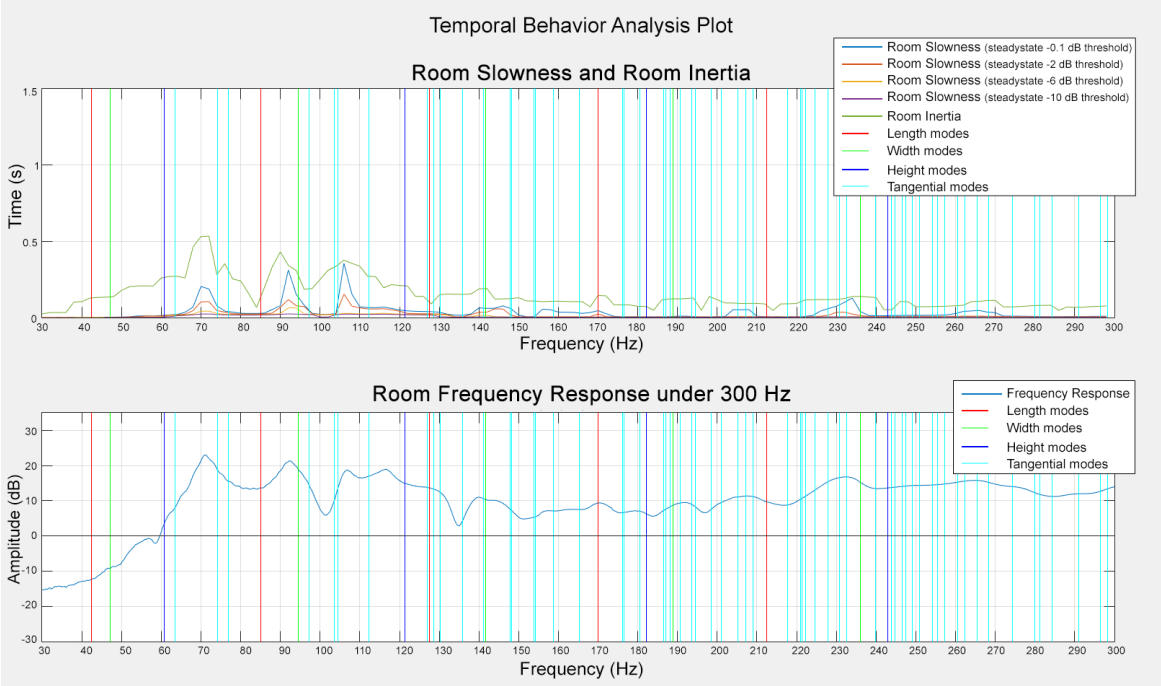


Figure 5.51: Room DrmReg - Temporal Behavior Analysis Plot

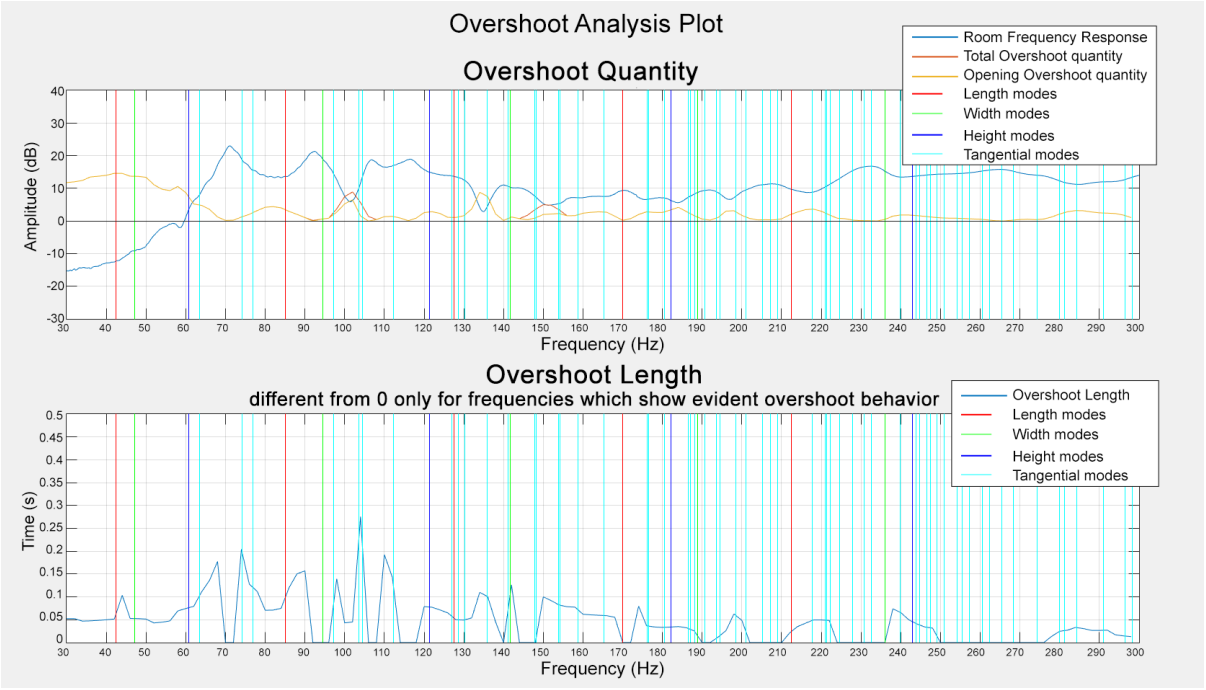


Figure 5.52: Room DrmReg - Overshoot Analysis Plot

5.12 Room ranks with Slowness, Decay Time, Inertia values

Table 5.3 shows all values of average Slowness, Decay and Inertia for all rooms between 30 and 300 Hz. These values were chosen in order to focus on the low frequency area where room modes are most significative. The area under 30 Hz was ignored since it was below the cutoff frequency of the speaker used to produce the sine sweep during the impulse response measurement, in order to avoid artifacts. The average Room Slowness, for this calculation, was as already stated, computed with the Room Slowness variant with 10 dB as threshold value. This variant was chosen, as already explained, as it seemed the most significant to represent the overall behavior of the Response Envelopes in their rise time, disregarding the final, possibly slow part of the slope and small fluctuations. Since it describes the first part of the slope of the Response Envelopes, it is mostly related to the initial and most significative contribution to the rise time of the envelope. These values were computed as the medium value of the vectors containing the Slowness, Decay or Inertia values for that specific room.

Room	Room Slowness [s]	Room Decay Time [s]	Room Inertia [s]
Room CN	0.0077	0.4779	0.1880
Room PRZ	0.0047	0.1682	0.1659
Room SGR	0.0054	0.4416	0.0935
Room GD	0.0054	0.5257	0.1951
Room SNT	0.0158	0.6497	0.3692
Room DrmA	0.0035	0.3779	0.1507
Room DrmB	0.0030	0.4511	0.1601
Room DrmReg	0.0060	0.2638	0.1486

Table 5.3: Values of Room Slowness, Decay Time, Inertia for all rooms

This computation was done, as already stated, after the first psychoacoustic test. Some of the most interesting rooms were chosen to test the psychoacoustic perception of these parameters, and if the perceived precision and definition of sounds was correlated to these metrics. These rooms are CN, PRZ, SGR, SNT and they rank slightly diffrently, depending on which parameter was chosen.

Rank with Room Slowness parameter (from lowest to highest):

- Room PRZ (0.0047 s)
- Room SGR (0.0054 s)
- Room CN (0.0077 s)
- Room SNT (0.0158 s)

Rank with Room Decay Time parameter (from lowest to highest):

- Room PRZ (0.1682 s)
- Room SGR (0.4416 s)

- Room CN (0.4779 s)
- Room SNT (0.6497 s)

Rank with Room Inertia parameter (from lowest to highest):

- Room SGR (0.0935 s)
- Room PRZ (0.1659 s)
- Room CN (0.1880 s)
- Room SNT (0.3692 s)

As it is possible to see, ranks generated with Slowness and Decay Time are the same, while the rank generated by Inertia features two rooms with switched positions. Some questions in the second psychoacoustic test will be aimed at evaluating wether these metrics can describe the psychoacoustic perception of the loss of precision and definition in sounds when convolved with room impulse responses.

5.13 General conclusions on AQT Room Analysis

As the previous subsections show, these are the main results of the AQT analysis with bass envelopes:

- Two main behavior and areas of interest emerge from the analysis: the overshoot behavior, on valleys, which is hypothized to act on the perceived volume of those frequencies when varying the note duration, and the slowness when reaching the and decaying from the steady state, on frequency response's peaks, which is hypothized to act on the perceived definition and precision of those frequencies.
- For long notes, more frequencies' AQT Response Envelopes reach their steady state
- Peaks in the frequency response have a Response Envelope which tends to reach their steady-state slower than other frequencies and to not have overshoot behavior. This is expected in a damped resonant system.
- Valleys in the frequency response have a Response Envelope which tends to show overshoot behavior and, after that, to reach their steady state very fastly. This is expected in a damped resonant system.
- In general, lower frequencies peaks have a Response Envelope which tends to be slower than higher frequencies peaks in reaching their steady state
- Room Slowness and Room Inertia seem to be correlated and feature high values on room modes, while they feature low values on valleys.
- From the Overshoot Advanced Analysis, it appears that the presence of a valley is correlated to the presence of overshoot behavior.

It is clear that the temporal behavior is very different for peaks and valleys of the frequency response. Peaks are slow in reaching their steadystate (for shorter notes they sometimes fail to reach it), do not show overshoot behavior and feature longer decay times as the next sections will show.

Some further things to keep in mind:

- Some rooms may have irregular shapes, in this case the frequency response should be considered more important than the computed resonance modes in the graph analysis.
- If the frequency that shows the higher overshoot or the slower behavior is slightly moved from the FFT peak or valley frequency, it could mean that some modes are interfering with each other.
- On the frequencies where H(f) and H(overshoot) are the same, that does not necessarily mean that the system has reached its steady state, especially for short test tones. In fact, the steady state was defined as the value of the envelope at the end of the test sound for the selected note duration.
- The longer the test tones, the more confidently it is possible to state that frequencies reach their actual, theoretical steady state value. The 550 msec test tones can be considered to reach their steady state for all frequencies. In this case, the AQT "steady state" curve is the same as the theoretical Fourier transform (which is the first subplot of the AQT results).
- On the frequencies where H(f) and H(overshoot) are different, this means that the steady state is actually lower than the overshoot, and the amount of this effect can be seen by the Overshoot Analysis plot which is important for short stimuli.
- The Overshoot Response is hypothized to be better than the classic Frequency Response when evaluating short sounds volume perception. In particular, the Overshoot Response developed with 150 milliseconds bursts should be used as it incorporates the effect of rooms with high "Slowness" and "Inertia" values, by featuring a lower value on the frequency response's peaks whose AQT Response Envelope does not reach its theoretical steady state value for really short notes.

Chapter 6

Psychoacoustic Tests

6.1 Why Psychoacoustic Tests

The aim of this thesis is, in the first place, to analyze how much people and musician are able to perceive the effects of low frequency room modes. Therefore, it is inevitable that listening tests are used to assess this. Listening tests have been always used to evaluate psychoacoustic effects such as masking, loudness, pitch perception, spatial perception.

These tests are very different from audiometric tests: audiometric tests aim at evaluating the tester's ability to discern sounds. Psychoacoustic tests aim at describing the sound with the tester's subjective opinion. The word "subjective" is very important: there is no right or wrong answer, especially in these two tests. Testers were aware of this and were invited to answer with what they heard, without worrying of what could be the "expected" answer. This kind of "non-judgemental" condition is very important to avoid possible sources of biasing.

In order to further put testers at ease, all audio files could be listened multiple times until the tester was sure about his or her answer.

The data will have to be processed with statistics, because that it is the only way to understand which effects have been heard by most people.

6.2 Test Devices

This section describes the devices used to perform the psychoacoustic tests. The same computer, DAW, listening levels and heaphones were used for all tests and all testers, to avoid the risk of introducing variation in the results. All tests were performed in quite places which did not have disturbing background noise.

6.2.1 Playback System

The playback system consisted of a 2010 MacBook Pro 17", 2.8 gHz Intel Core 2 Duo, 8 gb 1067 Mhz DDR3 with OS 10.9.5, using the digital audio workstation Avid Pro Tools 11.

The channel faders and master fader in Pro Tools were set at the default 0 dB value without any insert or plugins, and the audio files were developed in order to not "clip", which means, to not go over the digital 0 dB value. The output volume of the MacBook Pro was set at 9 over 16. The author had no way to exactly

measure the sound pressure level that reached the ears of the testers, so this value was chosen taking into account the Weber's Law: smaller differences are heard more clearly when the listening volume is lower. Some testers asked if the volume could be raised, but the listening conditions had to be equal for everyone, so the request was denied.

6.2.2 Headphones

The headphones that were used in both psychoacoustic tests are AKG K530 LTD headphones, which are mid-level, circumaural (completely covering the ear), semi-open (not completely isolated by the outside environment) headphones with a good low-frequency response. An official frequency response of these headphones was not present on the company's website. An unofficial frequency response and waterfall plot was instead found on [2]. This is a measurement performed on a specific headphone set, therefore it may differ slightly from the one used in these tests. Despite having a non completely idea frequency response or temporal behavior, these effects apply for all sounds that are reproduced through the headphones.

Fig. 6.1 and 6.2 show the frequency response and waterfall plot of the headphones as measured by [2]. Unfortunately, the waterfall plot is missing the frequencies below 200 Hz, which are the ones most significative for this research.

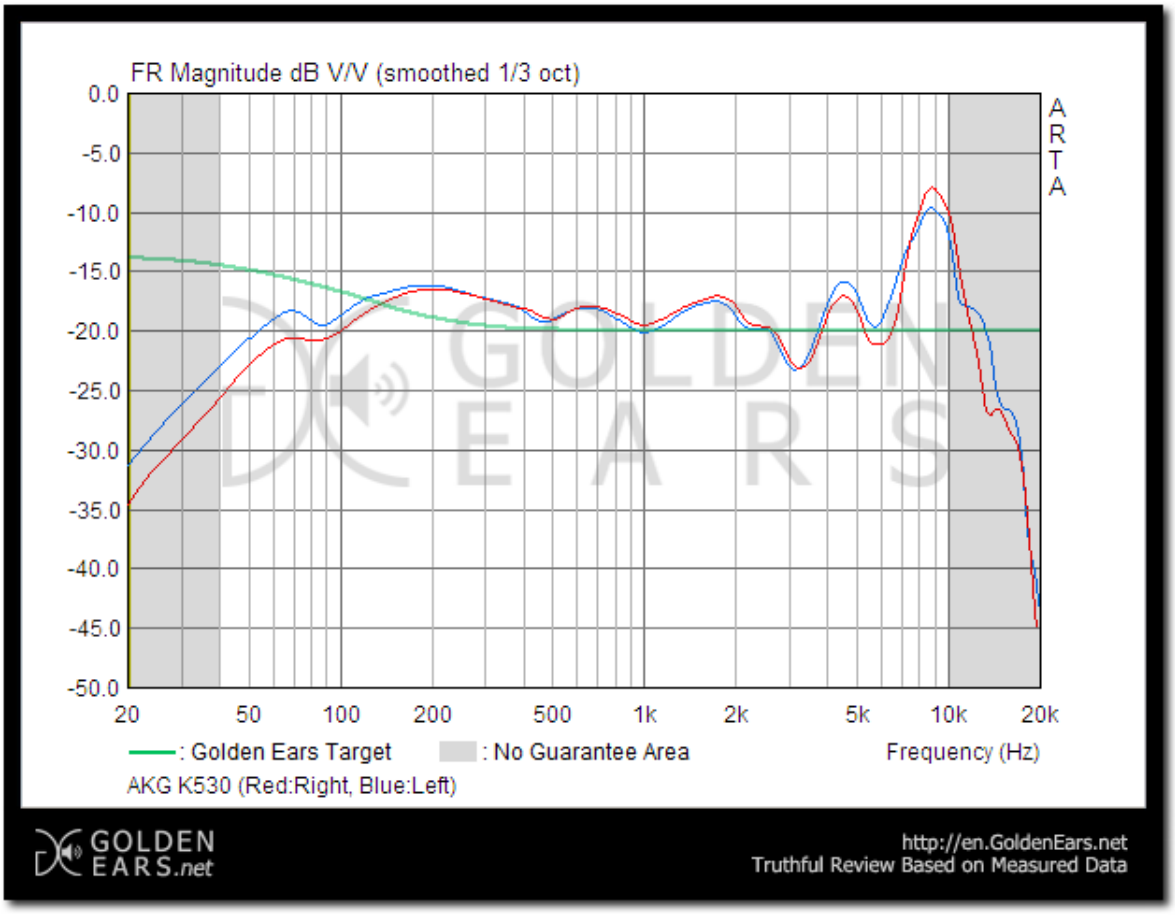


Figure 6.1: Headphones Frequency Response (unofficial)

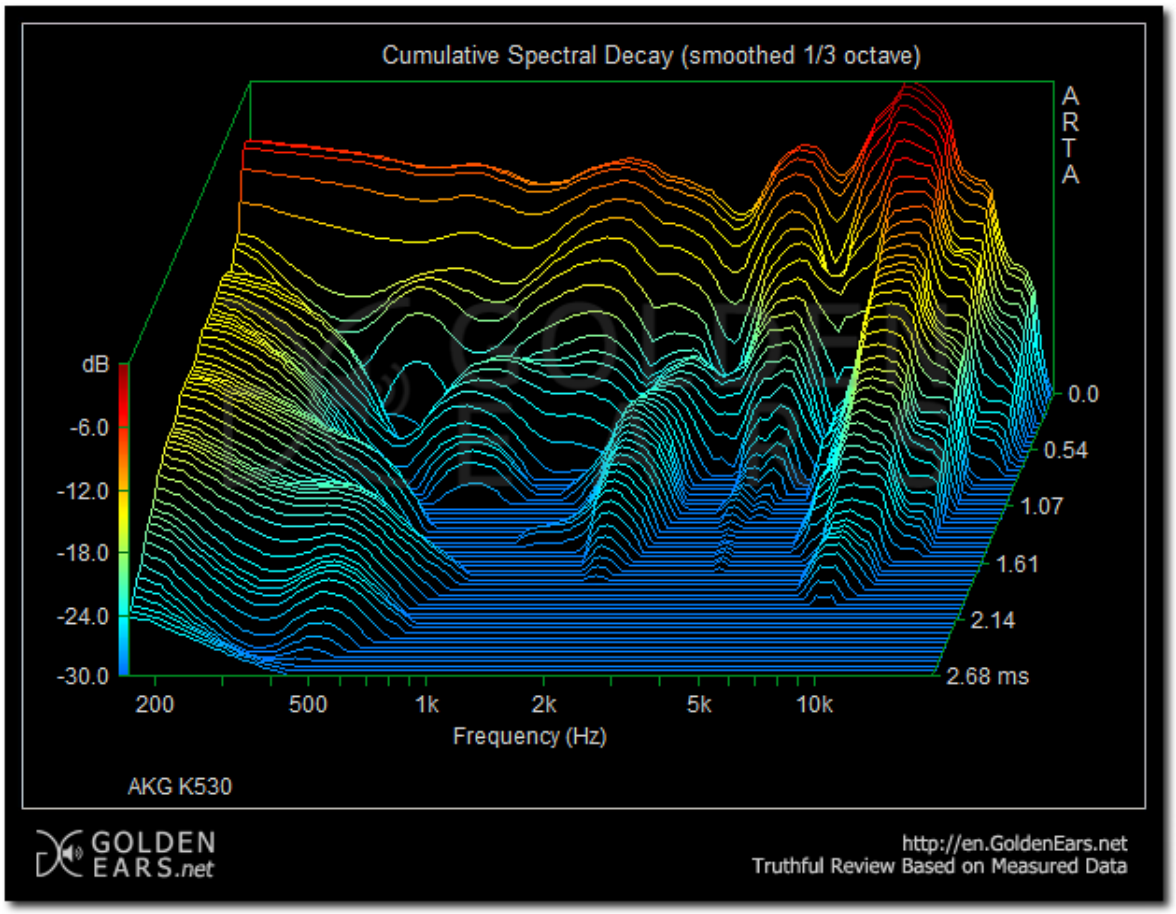


Figure 6.2: Headphones Waterfall Plot (unofficial)

6.3 Creation of the Test Files

The test files were developed in order to highlight some specific behavior, and the questions asked something about it. As an example, if the user could hear any difference between sounds and, if so, to quantify or describe it.

The author analyzed each Room Impulse Response with the developed AQT script, located frequencies which showed the problematic behavior caused by a certain effect, and developed the test audio files on those frequencies.

The effects that the author looked for, in order to develop the test questions and files, were: Overshoot behavior at a specific frequency, Room Slowness and Inertia at specific frequencies, difference between overshoot value and steady state value at a specific frequencies, frequencies which had the same overshoot behavior and different steady state value. These effects were tested with different sounds (kick sounds and bass sounds in the first test, kick sounds, bass sounds and musical excerpts in the second test) tuned on those frequencies and with different lengths for bass sounds (150, 250 and 550 milliseconds exactly like the duration of the AQT test signals) to assess the importance of the duration over the perception of a specific parameter.

It is important to remember that, in these tests, no variable is controllable since the Room Impulse Responses were real room ones. Therefore, it is not possible to change only one parameter (as an example, a specific frequency's decay time or overshoot value).

The test files were developed in Steinberg Cubase 8 using third-part virtual

instruments and sample libraries, as described in the following sections.

After the sounds were created, some of them (according to the questions of each test) were convolved with the Room Impulse Responses of the rooms where the sound's frequency showed some particular behavior. The convolution was performed in Matlab™ by a script made by the author. Some of them were processed in special ways in Matlab™ (as an example, a few sounds were cut in order to remove the overshoot section). This processing will be discussed during the description of each test's questions.

All files were uncompressed, 44.1 kHz, 24 bit, mono, .wav files. The choice of using mono sounds was made in order to avoid the possibility of introducing other unknown psychoacoustic effects, but repeating and developing these experiments with stereo sounds could be an interesting direction for further researches.

6.3.1 Kick Sounds

Kick sounds were created with "Kick - Drum Synthesizer Nicky Romero edition"™, a virtual instrument capable of generating percussive sounds with the fundamental frequency specified by the user. A midi note on a midi track triggered the input of the VST. The midi note does not influence the note generated by the program, which is instead specified in its settings. The sounds were then rendered to audio (wav, 44.1 kHz, 24 bit, mono files).

The parameters were set as shown in Fig. 6.3 As the figure shows, the envelope is made of four points that could be controlled in time and frequency. The first one was fixed at 20 khz. The third and fourth one were set at the note corresponding to the problematic frequency to be tested, and the second point was set exactly two octaves above that note in order to simulate the timbre of kick sounds generally used in music production.

Figure 6.3: Kick Sounds Settings

For different kick sounds, only the three points corresponding to the fundamental

frequency, and the one two octaves above, were modified, while all other values (sustain, click type and so on) were kept the same.

6.3.2 Bass Sounds

Bass sounds have been generated with the library Sample Tank 2.0™ by IK Multimedia. The chosen bass sound was J Bass Picked, which is a classic bass instrument played with a pick. The author chose to use a picked bass library because of the faster attack, that should excite more the overshoot behavior. The sound has been low-passed at 1000 hz, in order to remove some of the hi-end "rattle" caused by the pick attack that could, in the author's opinion, distract from the low frequency behavior on which the tests were focused on.

As already stated, problematic frequencies were spotted in each room by inspection of the AQT EFT plot. The conversion between the frequency number in Hertz and the note to be played has been looked up in [3]. Midi notes of length 150, 250 and 550 ms were generated at those notes, triggering the sound library. Sounds were rendered to audio and exported as 44.1 kHz, 24 bit mono .wav files.

6.3.3 Pure Tones

One question in test two featured a comparison between pure tones convolved with a Room Impulse Response. The pure tones were generated in Matlab™ with a script written by the author, with length of 0.55 seconds. They were convolved with the Room Impulse Responses in Matlab™ and finally rendered as 44.1 kHz, 24 bit, mono, .wav files.

6.3.4 Music Excerpts

In the second test, music excerpts were developed in order to test if the effects of room slowness and other parameters were perceived not only on single sounds but also with musical pieces. These were short (around 20 seconds) compositions with the same bass sound and kick sound as the other questions, and other sounds to complete the composition ("clap" hits, a piano, and a synth pad). The chosen BPM was 120.00, so that quarter notes last 0.5 seconds and sixteenth notes last 0.12 seconds (similar to 0.55 and 0.15 seconds of the test sounds).

A different music excerpt was developed for each of the room under test. In each one, the kick was tuned at the lowest resonant frequency of that room's frequency response, and the bass notes were all hitting the major peaks and valleys of the frequency response. For three of the four rooms the excerpts were developed for, these notes luckily were all included in the B flat major tonality, with allowed notes: B flat, C, D, D sharp, F, G, A. In these three clips, piano and synth pad parts were equal, to avoid distracting the tester. The fourth room, SNT, featured problematic frequencies which did not fit the previous tonality, so a new one had to be used. In this case, allowed notes were: A, B, C sharp, D, E, F sharp, G sharp. This was not perceived as a problem by any of the testers, as they were also instructed to evaluate the perceived parameters such as quality and definition without evaluating the musical content of the excerpt.

These files were also rendered as 44.1 kHz, 24 bit, mono, .wav files and were convolved with their relative Room Impulse Response in Matlab™.

6.3.5 Kick and Bass Spectral Content and expected impact on the test

Kick and Bass have a very different spectral content. In fact, bass notes are harmonic, and they are mostly made of a fundamental frequency and its upper harmonics, which are spectral lines centered on multiple frequencies of its fundamental. On the other hand, Kick sounds are not tonal sounds, therefore their spectrum is more complex and wide, having also non-harmonic components.

As far as the test goes, for comparisons of two sounds where only a single frequency has a different behavior in both sounds (as an example, the first and second questions of the second test), Kick sounds are expected to mask this behavior more than bass sounds, having a more dense spectral content whose frequencies could mask the behavior of a single one. Bass sounds, having a spectral content made mostly by harmonics, are expected to be more revealing of such effects, for the opposite reason.

6.4 Testers

Thirty testers were asked to take part in each one of the two psychoacoustic tests. Testers were chosen among musicians, producers and expert listeners. In particular, the author, alongside L.Rizzi and G.Ghelfi from "Suono E Vita", chose to interview as many drummers, bass players and producers as possible. This decision was made since these types of musicians are expected to be used to evaluate low frequency content in audio files and music. On the other side, a choice not to include people who didn't have a solid background as music listeners was made, in order to obtain a result that was meaningful for expert listeners. However, developing the same tests with non expert listeners could be an interesting area of research, in order to complete this analysis and study the difference between expert and non-expert listeners on these particular effects.

The first test had 17 testers who either played drums, bass or were producers. The second test had, instead, 16 such testers.

Among testers, there were expert listeners who did not play an instrument, and musicians such as drum players, bass players, violin players, piano players, electric and acoustic guitar players, singers, composers, producers.

Average testers age was 26.03 years in the first test and 25.63 in the second one. More detailed testers statistics will be shown in the results of each of the two psychoacoustic tests.

6.5 Possible Bias Sources

Psychoacoustic tests have to be prepared knowing that testers can be subject to biases, which are subjective factors affecting the results of a listening test. Differently from random errors, which are easy to recognize and can be taken care of with statistical tools, systematic errors manifest themselves as a consistent shift in the data, and systematic biases are difficult to get rid of, even with statistical tools ([47]).

The best way to deal with biases is to develop the tests in order to minimize them, eventually avoiding or minimizing errors when interpreting the results and limiting the error propagation when using the results as a basis for a new research.

This starts with a careful selection of test sounds, using audio files that are representative of the problem, that are preferably short and consistent (with no big changes in content in order to avoid the tester from remembering only the latest part of the latest sample and altering therefore his/her judgements). The sounds chosen for these tests are all short, and consistent by definition (most of them are just a single note). The listening environment varied with every subject, as tests were performed in the tester's houses or offices. This might introduce some bias in the results, but all environments were silent, and the fact that headphones were used should limit the influence of this variable to a negligible impact.

Also, there were lots of similar questions. In order to avoid the tester seeing a "trend" in their answers, some of the question's audio files were inverted (as an example, in some question the non-convolved sound was played before the convolved sound, other times it was the opposite).

6.5.1 Classic Bias sources

Each specific listener has his/her own experience and training with listening to music, therefore each one could give different answers to the test's questions. For this application, hearing impairment was not considered since test sounds were all focused at low frequencies. However, to the best of knowledge of the author and of the testers, no one had hearing impairment problems.

As extensively explained in [47] (which the reader is invited to read to grasp a better understanding of the subject), people evaluate audio content using sensory judgements, but also affective judgements. Affective judgements can be biased by the branding, appearance and price of the equipment. However, these variables were the same within all testers and all tests.

Biases can also arise from emotion (specific reaction to a stimulus) and mood (general feeling). In order to avoid upsetting or boring the testers, both tests could be finished quite fastly (around 15 mins each). In order to avoid letting the tester feel discouraged (and also to improve the quality of the restults), he/she was allowed to listen back to each sound until he was sure of his answer and willing to go to the next question.

6.5.2 Response Mapping Bias and Choice of scales for the second test

Questions which are answered on a scale can be source of lots of different bias problems, known as Response Mapping Bias, and arise from the fact that the mapping process of internal judgements onto a scale is not straightforward. A few of these are:

- Stimulus Spacing Bias: tendency to mark points on the scale in a non linear way with respect to the stimuli
- Stimulus Frequency Bias: tendency to evaluate slightly differently stimuli that are in reality the same

- Contraction Bias: tendency to avoid using the scale extremes when assigning judgements
- Centering Bias: having a different center reference point when assigning judgements
- Range Equalizing Bias: tendency to use the same range in the scale regardless of the size of the range of the stimuli
- Bias due to perceptually nonlinear scale: arises from the labeling of areas of the scales and different translations and interpretations of such labels

[47] provides a thorough explanation of all these possible bias sources.

In [47], an interesting discussion has been made regarding the use and validity of scales with labels. In particular, it is suggested that, in order to eliminate the nonlinearity problem, labels should not be used. However, labels are needed in some applications, so an alternative method could be to use only two labels at the scale's extremes. A test was conducted to see if labels had an impact, and, contrary to what was expected, the difference in results of questions that were answered with labeled and unlabeled scales was negligible. Nielsen and Meilgaard, instead, proposed to use labels which were only of a descriptive (sensory) nature.

Anchoring techniques have been suggested as ways to reduce some possible bias sources, since, with this technique, it is possible to partially "calibrate" the scale using a precise stimulus. Anchors are audio files that are used to define a point on the scale.

Some questions in the second test needed to be answered with a numeric value, in order to be later processed with ANOVA techniques. Therefore, two particular scales were developed by the author in order to deal with most of the possible aforementioned bias sources.

The first question was aimed at evaluating the perceived definition and precision of a sound after its convolution with a Room Impulse Response, to evaluate the effect of the room's Slowness and Inertia values. In order to perform ANOVA and statistical analysis, a numbered scale had to be used. Instead of asking the tester to answer with a value on a numeric scale, a graphic scale was developed with a color gradient from dark grey to light grey. This gradient allowed the scale to be perceived as a continous one, and this should minimize the tendency of answering with the most common numbers (like 50, 75, 100). Small dents were present on one side of the scale, and they stood for numbers from 1 to 100.

Since there were many sounds to be compared, an anchor was necessary to define the "reference" value of precision and definition for each comparison. The same sound (non-convolved) as the one that had to be valued (convolved), was used in each comparison as the anchor point. The testers were instructed to think of this sound as if it had "maximum precision", so, thinking of it as if it was on the far right, maximum end of the scale. Labels had to be used to quantify and describe the loss of definition/precision with respect to the anchor sounds. As one of the suggestions in [47], the author chose to use descriptive labels, such as "no degradation" (maximum value, attributed to the anchor), "light degradation", "medium degradation", "strong degradation", "extreme degradation". All labels were in italian, as all testers were italian. The tester heard first the anchor sound, then the test sound, and was asked to evaluate the precision and definition of the second sound

with respect to the first one (anchor) with a vertical line on the graphic scale. The author then translated each vertical line with the numeric value on the scale from 1 to 100 and processed this data statistically.

Anchoring was needed not only to correctly and consistently evaluate the second sound with respect to a reference, but also to evaluate all convolved sounds, by providing a stable reference (since all non-convolved sounds come from the same plugin and sound library) that is heard by the tester very often, avoiding him/her to misjudge sounds because of the last sounds he or she has heard that may obfuscate his perception and judgement abilities. A lower anchor was not used for two reasons: first, it would have made the test a lot slower, with the risk of boring and confusing the testers. Second, it would have been really hard to define a "minimum" in this scale. As an example, a tester who was also a developer of audio application, thought of the lowest end of the scale (maximum degradation) as digital ones and zeros, whereas other musicians often used the lowest end of the scale for sounds that were still comprehensible.

This scale is shown in Fig. 6.4.

Figure 6.4: Scale used in the second test (1)

The second question was aimed at rating a "global quality" of the room under analysis. This was a completely subjective question and the tester was invited to answer freely, taking into account factors such as definition, type of sound and his/her own personal preference. Since this question is more subjective, only two labels (of perceptive nature) were used at the ends of the scale: "low preference" and "high preference". Upper anchoring was also used, in the same way as the previous question. The anchor sound was exactly the sound under test, except it was not convolved with any Room Impulse Responses.

This scale is shown in Fig. 6.5.

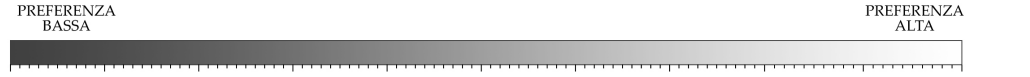


Figure 6.5: Scale used in the second test (2)

6.6 First Test

The first psychoacoustic test was performed in August, 2015. It lasted on average 15 minutes for each listener. Thirty testers had to listen to test sounds through headphones and answer on paper to simple questions. They were allowed to listen back to the test sounds how many times they wanted before answering. Playback level, devices, headphones were always the same.

Most of the questions compared two sounds and asked a simple question regarding the difference in perceived precision/definition (the validity of these words

will be discussed in the later sections). Most of the times, the comparisons were made between a non-convolved sound, and the same sound convolved with the Room Impulse Response of a room in which the sound's fundamental frequency showed a particularly interesting behavior. Different questions were also made, as will be discussed in the following sections.

6.6.1 Goals of the Test

The aim of this test was to assess the psychoacoustic effect of some of the phenomena that emerged during the AQT analysis. In particular:

Effects of the speed of the room to reach steady state

The test aimed at evaluating if the "Slowness" of each problematic frequency's Response Envelope could actually be perceived by the testers. Also,to verify if the same variable has an impact on the perceived precision and definition of the convolved sound. This was tested for different sounds (bass, kick) and different durations (long bass note, short bass note) in different rooms (room whose modes' AQT Response Envelope reached steady state for short bursts, and room whose modes' AQT Response Envelope did not).

Overshoot Response importance for short sounds

The test aimed at evaluating if the overshoot response could be more significative than the classic frequency response for short sounds when assessing the note's perceived level, and how much this effect would change for longer notes.

Note Duration and Room Mode presence on pitch perception

The test aimed at evaluating if the same notes with short duration were still perceived as having the same pitch, and if the presence of room modes alters the ability of correctly assessing the pitch.

Room modes effects on perceived volume

The test aimed at evaluating if the presence of a room mode would "energize" the note making it seem louder than a note which wasn't centered on a room mode, even if the two nominal levels were the same. This idea was introduced by M. Ferroni, L. Rizzi and G.Ghelfi in M. Ferroni's thesis work, and needed further testing.

6.6.2 Test Structure and Questions

The list of questions, description of test sounds and goals of each question will be now described.

Audio files naming scheme is as follows: TypeOfSound_Duration_Frequency_ConvolvedRoom. As an example, Bass_015_38_PRZ is a 0.15 second bass note at 38 Hz convolved with room PRZ's impulse response. If the last field is left empty, it means that the note was not convolved. Kick notes have duration 015 in their naming by default.

Of course, the name of the files were hidden to the testers to avoid bias. The testers could only read the questions and listen to the files, but the file names are

written here so that the reader can understand better why the questions have been made.

1. Kick_015_44_CN - Kick_015_38_PRZ

   (a) Which sound is the most precise?

   (b) Which sound is the most resonant?

   Goals: to assess if testers are able to hear which sound has greater precision when comparing two sounds convolved with different room Impulse Responses; to assess if "precise" and "resonant" were descriptive of the phenomena under test; to assess if they were considered as opposite words.

   In particular, the compared sounds were short bass notes at 44 Hz (main peak of room CN) convolved with room CN (whose main peak's AQT Response Envelope is slow in reaching its steady state value), and at 38 Hz (main peak of room PRZ) convolved with room PRZ (whose main peak's AQT Response Envelope is fast in reaching its steady state value).

   Expected outcome: the author expects the words to be the opposite of one another, and that the testers will correctly evaluate the most precise sound as the one convolved with room PRZ (where the peak's AQT Response Envelope reaches fastly its steady state) and, as the most resonant, the one convolved with room CN (opposite behavior).

2. Bass_015_38_PRZ - Bass_015_44_CN

   (a) Which sound is the most precise?

   (b) Which sound is the most resonant?

   Goals: same as question 1, but tested on short bass notes.

   Expected outcome: same as question 1.

3. Kick_015_44 - Kick_015_44_CN

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

   Goals: to understand if testers are able to perceive the modification in precision and definition between the non-convolved, dry sound and the same sound convolved with the Impulse Response of a room in which that frequency's AQT Response Envelope is slow in reaching its steady state value.

   Expected outcome: the author expects that most testers will be able to correctly evaluate the convolved sound as the one with less precision, since it is convolved with a "slow" room (CN).

4. Kick_015_38 - Kick_015_38_PRZ

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

Goals: to understand if testers are able to perceive the modification in precision and definition between the non-convolved, dry sound and the same sound convolved with the Impulse Response of a room in which that frequency's AQT Response Envelope is fast in reaching its steady state value.

Expected outcome: the author expects that some testers will be able to correctly evaluate the convolved sound as the one with less precision, since it is convolved with a "fast" room (PRZ). However, fewer people are expected to notice this behavior with respect to the previous question, because of the fast speed of most frequencies' AQT Response Envelope in reaching their steady state value.

5. Kick_015_44 - Kick_015_44_CN - Kick_015_38 - Kick_015_38_PRZ

   (a) Is there a bigger precision degradation among sounds of the first couple (from the first to second sound) or the second couple (from the third to the fourth sound) ?

   Goals: to understand if the tester perceived that a room whose lowest mode's AQT Response Envelope is really slow in reaching its steady state degrades the perceived precision more with respect to a room whose lowest mode's response envelope is faster.

   Expected outcome: the author expects that the testers will be able to correctly state that the higher precision loss happens in the first couple ("slow" room).

6. Bass_015_44_CN - Bass_015_44

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

   Goals: same as question 3, but tested with short bass notes.

   Expected outcome: same as question 3.

7. Bass_015_38 - Bass_015_38_PRZ

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

   Goals: same as question 4, but tested with short bass notes.

   Expected outcome: same as question 4.

8. Bass_015_44 - Bass_015_44_CN - Bass_015_38 - Bass_015_38_PRZ

   (a) Is there a bigger precision degradation among sounds of the first couple (from the first to second sound) or the second couple (from the third to the fourth sound) ?

   Goals: same as question 5, but tested with short bass notes.

   Expected outcome: same as question 5.

9. Bass_055_44 - Bass_055_44_CN

(a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

Goals: same as question 3, but tested with long bass notes.

Expected outcome: same as question 3.

10. Bass_055_38_PRZ - Bass_055_38

    (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

    Goals: same as question 4, but tested with long bass notes.

    Expected outcome: same as question 4.

11. Bass_055_44 - Bass_055_44_CN - Bass_055_38 - Bass_055_38_PRZ

    (a) Is there a bigger precision degradation among sounds of the first couple (from the first to second sound) or the second couple (from the third to the fourth sound) ?

    Goals: same as question 5, but tested with long bass notes.

    Expected outcome: same as question 5.

12. Bass_003_72 - Bass_003_72_PRZ (valley) - Bass_003_72_DrmREG (peak)

    (a) Is the pitch always the same?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if the presence of a valley or peak and the consequent change in tone alters the ability to correctly assess the note pitch for very short notes.

    Expected outcome: the author expects that, for very short notes, the change in timbre caused by the presence of frequency response valleys or peaks will be so strong to alter the pitch perception in some individuals.

13. Bass_015_72 - Bass_015_72_PRZ (valley) - Bass_015_72_DrmREG (peak)

    (a) Is the pitch always the same?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if the presence of a valley or peak and the consequent change in tone alters the ability to correctly assess the note pitch for short notes (slightly longer with respect to the previous question).

    Expected outcome: the author expects that most of the testers will perceive all notes having the same pitch.

14. Bass_055_72 - Bass_055_72_PRZ (valley) - Bass_055_72_DrmREG (peak)

    (a) Is the pitch always the same?

    (b) With how much confidence would you say this? (from 1 to 10)

Goals: to understand if the presence of a valley or peak and the consequent change in tone alters the ability to correctly assess the note pitch for long notes.

Expected outcome: the author expects that all of the testers will perceive all notes having the same pitch.

15. Kick_015_98_DrmB - Kick_015_49_DrmB - Kick_15_196_DrmB

    (a) Which sound is the most resonant?

    (b) Which sound is more precise/defined?

    Goals: to understand which note is perceived as more precise/defined and which as the most resonant, with notes having their fundamental frequency on the room's lowest mode, an octave lower, and an octave higher.

    Expected outcome: the author expects that the most resonant note will be the one tuned at 98 Hz (exactly over the frequency response's peak). The most precise/defined could be both the first and third note. However, because of equal loudness curves, the third sound (with fundamental at 196 Hz, will be perceived as more precise.

16. Bass_015_99_PRZ - Bass_015_111_PRZ

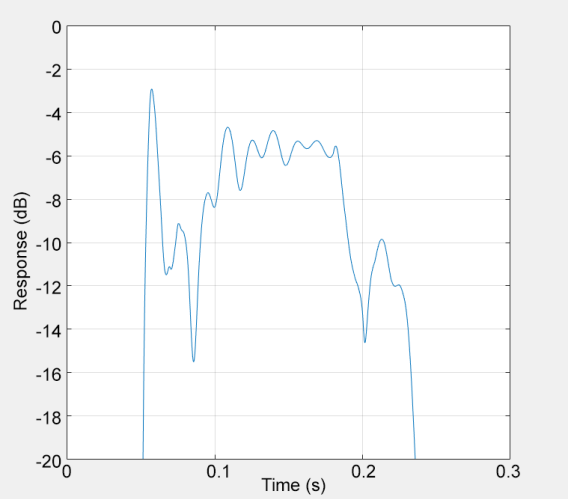
    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

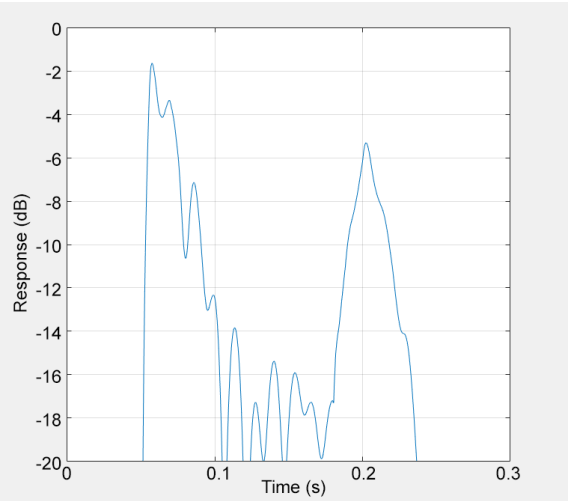
    Goals: to understand if notes having an almost identical Overshoot amplitude, but different steady state value, are perceived as having different volumes (on short notes).

    Room PRZ featured a peak at 99 Hz (its AQT Response Envelope is shown in fig. 6.6a) and a valley at 111 Hz (its AQT Response Envelope is shown in Fig. 6.6b) which had the same Overshoot level, but a very different steady state level.

    Expected outcome: the author expects that very few testers will be able to correctly state which note has the lower steady state value, since notes are very short; he expects that most testers will judge the sounds as having the same volume, and this would mean that the overshoot spike can be really important when assessing short notes' levels.



(a) Bass note at 99 Hz in Room PRZ



(b) Bass note at 111 Hz in Room PRZ

Figure 6.6: Bass notes at 99 and 111 Hz in room PRZ: similar overshoot behavior, different steady state level

17. Bass_025_99_PRZ_NoOvershoot - Bass_025_111_PRZ_NoOvershoot

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if notes having an almost identical Overshoot amplitude, but different steady state value, after removing the Overshoot portions (initial and final) and leaving only the steady state portions, are perceived as having different volumes (on short notes).

    Overshoots have been removed in Matlab™ after the convolution with the room impulse response, and short fade-in and fade-outs have been performed on the final sound.

    Expected outcome: the author expects that, since the overshoot is missing, testers should hear more clearly the steady state value and indicate which note is actually higher in volume.

18. Bass_015_111_PRZ - Bass_025_111_PRZ_NoOvershoot

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to assess if testers are able to hear the difference in volume between two sounds, where the first one has an overshoot and the second one is exactly the same as the first sound, but after removing both the initial and final overshoot portion. This answer compares the second sound of question 16 (with overshoots) with the second sound of question 17 (with no overshoots). The length of both sounds is the same, as the second one was created with a length of 250 ms and was later cut to a final length of 150 ms when its overshoot have been removed (in Matlab™).

    Expected outcome: the author expects that testers will identify the sound with overshoot as having higher volume, and the sound without overshoot as having lower volume.

19. Bass_015_246_SGR - Bass_015_280_SGR

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if notes having an almost identical Overshoot amplitude, but different steady state value, are perceived as having different volumes (on short bass notes).

    This room has a peak at 246 Hz and a valley at 280 Hz. Figs. 6.7 and 6.8 show the bass notes at both frequencies, before and after convolution with SGR Impulse Response, showing that the overshoot behavior is very similar, while the steady state level is different.

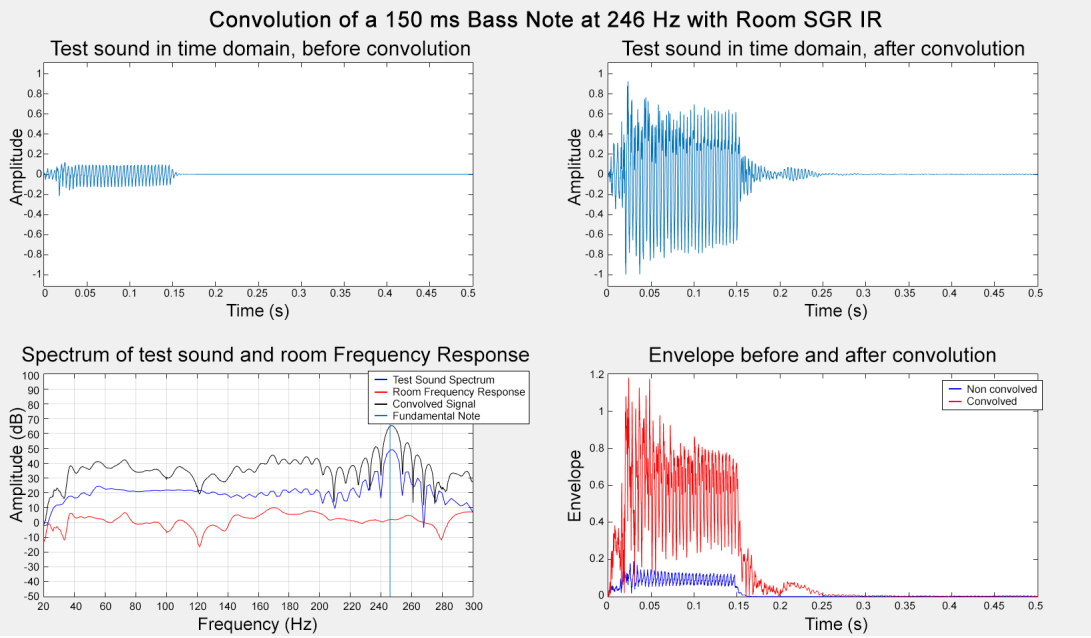


Figure 6.7: Bass note at 246 Hz in room SGR before and after convolution

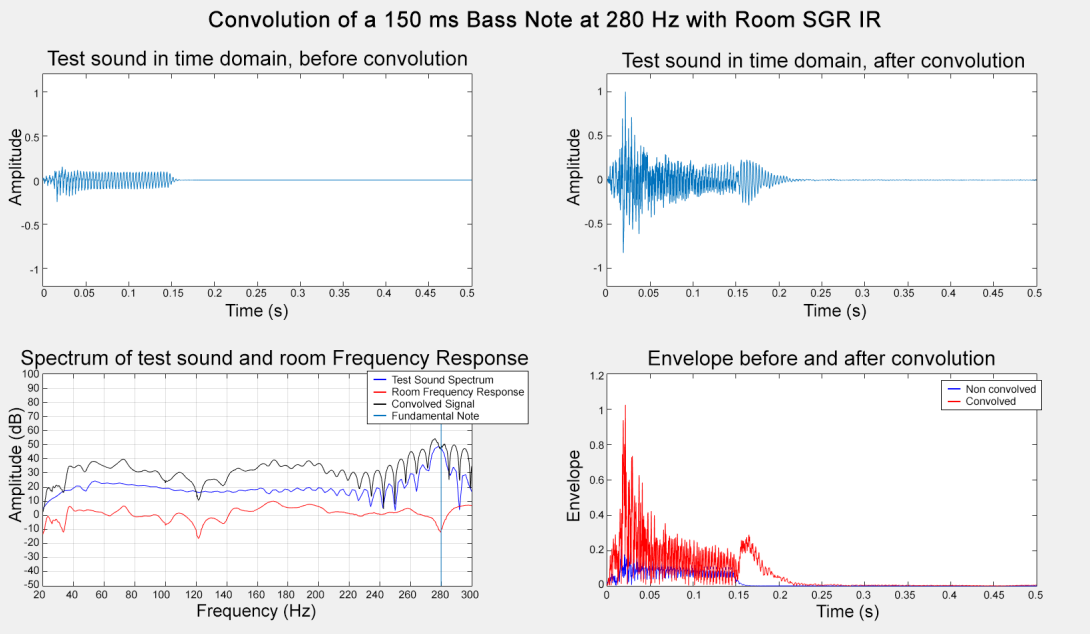


Figure 6.8: Bass note at 280 hz in room SGR

Expected outcome: the author expects that the note with lower steady state will be perceived as being quieter by some of the testers. Since the duration of the steady state is short, it is possible that some (or most) testers will not be able to hear a difference.

20. Bass_055_246_SGR - Bass_055_280_SGR

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if notes having an almost identical Overshoot amplitude, but different steady state value, are perceived as having different volumes (on long bass notes).

    The reader is reminded that, varying the length of the note, the overshoot behavior remains the same (both in amplitude and in duration) so, what varies is only the duration in which the note is in its steady-state level.

    Expected outcome: the author expects that the note with lower steady state will be perceived as being quieter by most testers, since the duration of the steady state part is longer than in the previous question.

    Note: questions 19 and 20 were really interesting, but the chosen frequencies were probably too high to be free from other factors that could impact the psychoacoustic perception (behavior of overlapping modes, interference, behavior of higher octaves of modes, equal loudness curves...). These frequencies where chosen because they showed the most similar overshoot behavior, while having the most different steady state level, among the examined critical frequencies in all rooms.

    Therefore, this and the previous question were remade in the second test on lower frequencies with similar overshoot and steady state behavior.

21. Bass_015_111_PRZ - Bass_025_111_PRZ_- Bass_055_111_PRZ

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if the duration of the notes has any impact on the perception of the volume, on a room in which that frequency's AQT Response Envelope reaches its steady state value fastly.

    Expected outcome: the author expects that the longest note will be heard as louder, since it is tuned at a resonant mode of the room. However, since room PRZ is "fast" (in the sense already explained), few testers are expected to give this answer.

22. Bass_015_44_CN - Bass_025_44_CN_- Bass_055_44_CN

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to understand if the duration of the notes has any impact on the perception of the volume, on a room in which that frequency's AQT Response Envelope reaches its steady state value slowly.

    Expected outcome: the author expects that the longest note will be heard as louder, since it is tuned at a resonant mode of the room whose AQT Response Envelope is really slow in rising towards its steady state value.

23. Bass_055_44 - Bass_055_44_CN

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Goals: to assess if the convolution with an impulse response of a frequency that is a room mode of that room, energizes the note letting the tester hear a higher volume even if the two sounds have been normalized with respect to their amplitude peak.

    Expected outcome: the author expects the convolved sound to be perceived as slightly louder. This idea comes from M. Ferroni's thesis work [37].

    For this question, the author has thought about different types of normalization. However, the one that was best for this application was normal "peak" normalization, that means, making the peaks of both audio files having the same amplitude. It is clear that a room mode (having an high amplitude) makes the note louder with respect to the non-convolved one (so, if the normalization was done with respect to the RMS value, the convolved one would still result louder), but is the note perceived as louder also when their amplitude peak is the same? This aims at further inspecting the concept of "energization" hypothized in M. Ferroni's thesis work.

6.7 Second Test

The second psychoacoustic test was performed in November, 2015. Before developing it, further research had been made. In particular, Temporal Behavior analysis was performed, leading to the definition of the parameters Room Slowness and Room Inertia and, as already explained, the ranking of four different particular rooms with respect to these parameters.

As the next section will show, the results of most "With how much confidence would you say this? (from 1 to 10)" questions of the first test was not really significative, as all values averaged between 7 and 8. Therefore, these questions were avoided in the second test.

6.7.1 Goals of the Test

The aim of this test was to expand the knowledge obtained by the first test, by asking more detailed questions about the same parameters and asking new questions about new parameters (Room Slowness and Inertia). Also, its aim was to confirm or develop a vocabulary of perceptive adjectives. In particular:

Confirm previous test's results

Some of the first test's questions were repeated with slightly different conditions in order to confirm their results, or to improve the conditions that would highlight the behavior under study (as an example, a question was repeated in a frequency range that would be more significative for our analysis).

Some new questions were made, assessing the same problems of the first test, but with different strategies with regards to the question type, in order to confirm the results of the first test.

Integrate previous test with musical excerpts and pure tones

While the first test featured only kick and bass sounds, this test employed also musical excerpts in order to evaluate if some room parameters were significative also when listening to musical pieces instead of notes played by themselves.

Also, a question asked a comparison between two pure tones, each one being convolved with an Impulse Response, in order to understand if a particular behavior was immediately recognizable when listening to sine waves, while instead being more difficult to hear when in presence of complex tones.

Connect room preference to acoustic parameters

As already stated, three room ranks were developed according to the metrics "Room Slowness", "Room Inertia", "Room Decay Time", ranking four rooms with interesting features in either their frequency response curve, temporal behavior or overshoot behavior. The author wanted to understand how these parameters were related to the subjective perception of the precision and definition degradation, and to the overall preference of a listener when assessing the listening quality of a room. Therefore, some questions were developed in order to let the tester evaluate each audio file (therefore, each room) on a scale from 1 to 100.

The aim was to see which one among the three ranks would be closer to the final ranking resulting by the test, in order to understand which, among Slowness, Inertia and Decay Time was the most impactful on the perceived loss of precision and overall preference and if these parameters alone would be able to describe both the precision loss and overall preference behavior or if something else should be taken into account.

Confirm or develop a vocabulary of perceptive adjectives

During the first, test, some subjects asked clarification about the meaning of the word "definition" and "precision" with regards to the test sounds. The reader is reminded that these two words have been chosed in M. Ferroni's thesis work, where he found the most adequate terms to describe some psychoacoustic effects by asking a pool of testers. While a rigorous definition of these words was not given, most testers were able to answer all questions yelding results that followed the author's expectation, implying that the inner meaning of the words was comprehended by the testers. However, in order to make sure there were no better words to describe such phenomena, some questions were developed asking testers to describe some sounds using words.

6.7.2 Test Structure and Questions

1. Kick_015_32_PRZ - Kick_015_38_PRZ

   (a) Which sound is the most precise?

   (b) Which sound is the most resonant?

   Goals: to understand if different values of room Slowness / Inertia on a single frequency impact the perceived precision of the sound; to confirm if precise and resonant could be used as opposite words.

   The reader is reminded that the Temporal Behavior Analysis block was introduced after the first test. As Fig. 5.33 shows, room PRZ has the highest frequency response peak at 38 Hz, (this was the frequency that was used in most of the first test's questions), however, 32 Hz shows higher Slowness / Inertia values, while being slightly lower in amplitude. After the temporal analysis, the importance of testing these frequency

   Expected outcome: the author expects that sound whose frequency has higher Slowness/Inertia values will be perceived as being less precise. If the effect is not heard, probably it is because higher frequencies mask this behavior, since it happens only on the sound fundamental.

2. Bass_055_32_PRZ - Bass_055_38_PRZ

   (a) Which sound is the most precise?

   (b) Which sound is the most resonant?

   Goals: same as question 1, but tested with long bass notes.

   Expected outcome: same as question 1. However, in the author's opinion, more people would answer correctly since notes are longer and sustained (making it

possible for the listener to focus on the reverberant tail at the end of the note), and since the frequency spectrum of the bass, being made of different harmonics, would mask the fundamental frequency in a less significative manner with respect to the kick (whose spectrum is composed by even non-harmonic frequencies).

3. Kick_015_58_PRZ - Kick_015_32_PRZ

   (a) Which sound is the most precise?

   (b) Which sound is the most resonant?

   Goals: same as question 1.

   However, the first sound of this question was centered on a valley, therefore having lower Slowness and Inertia values.

   Expected outcome: same as question 1, however the author expects more testers to answer correctly, since the two sounds have an higher difference of Slowness and Inertia values with respect to the sounds of question 1.

4. Kick_015_52_SNT - Kick_015_44_CN - Kick_015_38_PRZ - Kick_015_38_SGR

   (a) Rank the four sounds from the most precise/defined to the least precise/defined.

   Goals: to understand if the final rank would be closer to the rank generated according to Slowness, Inertia or Decay Time values. Therefore, understanding which of these parameters has a bigger influence on the perception of precision and definition degradation.

   Expected outcome: the author expects that the resulting rank could be the same as the ones generated with Inertia or Decay Time values. In the author's opinion, final decays are more perceivable than the initial slowness of some frequencies, therefore he expects that the rank generated with Slowness values is not significant for this application.

5. Kick_015_38_SGR - Kick_015_32_SGR

   (a) Do both sounds have the same volume? if not, which one is the quietest?

   Goals: to understand if, for percussive short sounds, the overshoot is more significative than the steady state value when assessing levels.

   On room SGR, 32 Hz is the lowest point of a valley and 38 Hz is a frequency response peak. The AQT Response Envelopes of these frequencies have a very similar opening Overshoot behavior and a different steady state value (being a valley and a peak). Of course, the frequency at 38 Hz does not have an overshoot, being a peak, but its opening part has the same amplitude as the overshoot peak of the valley. This question and the following two could be considered as a remake of question 19 and 20 of Test 1.

   Expected outcome: the author expects that the sounds will be mostly perceived as having the same volume, since they are percussive short sounds with a similar initial level (normal AQT Response Envelope level for the peak frequency, Overshoot level for the valley frequency).

6. Bass_015_38_SGR - Bass_015_32_SGR

   (a) Do both sounds have the same volume? if not, which one is the quietest?

   Goals: same as question 5, but tested with short bass notes.

   Expected outcome: same as question 5.

7. Bass_055_32_SGR - Bass_055_38_SGR

   (a) Do both sounds have the same volume? if not, which one is the quietest?

   Goals: same as question 5, but tested with long bass notes.

   Expected outcome: the author expects that more testers will be able to correctly state which sound is the quietest (the one whose fundamental frequency is on the valley). In fact, since the sound is longer and most frequencies are able to reach their steady state, the theoretical Fourier Frequency Response curve should become more significative than the Overshoot Response.

8. Bass_015_38_SGR_noOvershoot - Bass_015_32_SGR_noOvershoot

   (a) Do both sounds have the same volume? if not, which one is the quietest?

   Goals: to assess if testers are able to answer correctly to question 5 when the opening part (containing the Overshoot in the sound corresponding to the frequency response's valley) have been removed in both sounds. This question is similar to, and developed in the same way as, question 17 of Test 1 in order to confirm its result.

   Expected outcome: the author expects that more testers (with reference to question 5) will be able to correctly state that the sound whose fundamental frequency is the valley, is the quietest, since removing overshoots exposes the actual steady state value which is different for the two sounds.

9. Bass_015_32_SGR - Bass_015_32_SGR_noOvershoot

   (a) Do both sounds have the same volume? if not, which one is the quietest?

   Goals: to assess if testers are able to hear the difference in volume between two sounds, where the first one has an overshoot and the second one is exactly the same as the first sound, but after removing both the initial and final overshoot portions. This question is similar to, and developed in the same way as, question 18 of Test 1 in order to confirm its result.

   Expected outcome: the author expects that most testers will state that the quietest sound is the one without Overshoot portions, since its steady state level is exposed.

10. PureTone_055_32_SGR - PureTone_055_38_SGR

    (a) Do both sounds have the same volume? if not, which one is the quietest?

Goals: to show that the Pure Tone analysis actually yelds perceptually correct results and that, if testers do not hear some behavior on a specific frequency, it is probably caused by masking in the frequency domain.

Since these sounds are pure tones, their envelope is exactly the AQT Response Envelope of their frequencies. Therefore, both sounds feature some initial and/or final overshoot behavior, that is clearly audible as a small bump, or click. Testers were instructed to evaluate both sounds as a whole, asking themselves which of the sound was overall louder and quieter.

Expected outcome: the author expects that most testers will answer correctly to this question, indicating as the quietest sound the one whose fundamental frequency is on the valley.

11. Bass_015_44 - Bass_015_44_CN

    Bass_015_36 - Bass_015_36_SGR

    Bass_015_38 - Bass_015_38_PRZ

    Bass_015_52 - Bass_015_52_SNT

    (a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.

    Goals: to assess the perceived loss of precision of short bass notes in rooms with different temporal behavior.

    Testers had to answer on the scale pictured in Fig. 6.4.

    Expected outcome: the author expects that the best rated rooms will be PRZ and SGR (since they have the lowest value of room Slowness, Inertia and Decay Times) while CN and SNT should be the lowest rated rooms (having higher values of such parameters).

12. Bass_055_38 - Bass_055_38_PRZ

    Bass_055_44 - Bass_055_44_CN

    Bass_055_52 - Bass_055_52_SNT

    Bass_055_36 - Bass_055_36_SGR

    (a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.

    Goals: to assess the perceived loss of precision of long bass notes in rooms with different temporal behavior.

    Testers had to answer on the scale pictured in Fig. 6.4.

    Expected outcome: same as question 11.

13. Kick_015_36 - Kick_015_36_SGR

    Kick_015_52 - Kick_015_52_SNT

    Kick_015_38- Kick_015_38_PRZ

    Kick_015_44 - Kick_015_44_CN

    (a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.

    Goals: to assess the perceived loss of precision of kick hits in rooms with different temporal behavior.

    Testers had to answer on the scale pictured in Fig. 6.4.

    Expected outcome: same as question 11.

14. Excerpt_44 - Excerpt_44_CN

    Excerpt_38 - Excerpt_38_PRZ

    Excerpt_52 - Excerpt_52_SNT

    Excerpt_36 - Excerpt_36_SGR

    (a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.

    Goals: to assess the perceived loss of precision of musical excerpts in rooms with different temporal behavior.

    Testers had to answer on the scale pictured in Fig. 6.4.

    In order to avoid testers to judge the musical content or their overall preference regarding sound quality, testers where reminded that they had to give an answer based only on that

    Expected outcome: same as question 11.

15. Excerpt_44_CN - Excerpt_38_PRZ - Excerpt_52_SNT - Excerpt_36_SGR

    (a) For each sound, draw a vertical line on the scale on the point that represents your subjective preference of that sounds' listening quality.

    Goals: to assess the subjective listening preference of the testers and to study its relations with respect to the parameters introduced in this research.

    Testers had to answer on the scale pictured in Fig. 6.5. Testers were instructed to answer to this question in a completely subjective way.

    Expected outcome: the author expects that the behavior should be the same as the previous question, with a small difference: since room PRZ features a huge valley in the frequency response domain, showing a lack of bass, (and

since excerpts are really long, these sounds are well-represented by the theoretical frequency response curve, as opposed to short sounds being probably more well-represented by the Overshoot response), some testers may prefer room SGR to room PRZ. In fact, while room PRZ has slightly lower Slowness and Inertia values, room SGR has a flatter frequency response. So, some subjects may prefer a slightly less precise temporal behavior in favor of a better balanced frequency content.

16. Bass_015_44 - Bass_015_44_- CN

    (a) For each sound, draw a vertical line on the scale on the point that represents your subjective preference of that sounds' listening quality.

    Goals: to understand it the words "precise", "precision", "defined", "definition" were good descriptor of the psychoacoustic losses that a sound undergoes after the convolution with the Impulse Response of a room.

    Expected outcome: the author expects those words to be the ones used more frequently, and that no new words (or just a few) will be introduced.

17. Bass_015_38_PRZ - Bass_015_44_- CN

    (a) For each sound, draw a vertical line on the scale on the point that represents your subjective preference of that sounds' listening quality.

    Goals: same as question 16, but tested with two already convolved sounds (the first one with a room with low Slowness and Inertia values, the second one with a room with low Slowness and Inertia values).

    Expected outcome: same as question 16.

After the end of all tests, test results were analyzed.

Chapter 7

Test Results

7.1 First Test results

After performing the first psychoacoustic test on thirty listeners, results were analyzed.

7.1.1 Tester's statistics

All testers were musicians or expert listeners. This choice was made in order to test the phenomena introduced earlier with people who were trained to listen for small details in music. Since this thesis focuses on low frequencies, the author tried to find as many testers as possible who were drummers, bass players or producers. As a matter of fact, these types of musicians are the one who are mostly used to critical listening in the low frequency area.

Number of testers: 30
Number of bass players, drummers or producers among testers: 17
Average age of testers: 26 years
Younger tester: 18 years
Older tester: 40 years
Male testers: 28
Female testers: 2
Average duration of the test: 14 min
Faster test: 9 min
Slower test: 20 min
People who experienced listening fatigue at the end of the test: 3
People who were annoyed / felt uncomfortable because of the many questions: 1
Notes:

- Some people asked to take the test with an higher volume. This request was denied in order to have the same listening conditions for all subjects, and to take advantage of Weber's Law principle that states that a small difference between stimuli is more perceptible when stimuli themselves have a small amplitude.

- Question 23 was performed for all testers at two different listening levels, because the author felt that it could impact the results: the normal listening

level that the rest of the test had, and a higher level (all settings were the same, besides the volume of the Mac Book Pro that was set at maximum).

- Some people made question asking to better define the terms "precise". This indicates that this word alone (found by M. Ferroni in his thesis work) might not be completely explicative of the phenomenon of degradation cause by the convolution with an impulse response. However, by stating that it could be considered as a synonym of "definite", testers seem to have understood the meaning, answering questions mostly in the expected way.

- Since all testers were italian and the test was performed in Italy, the test language was italian. The words "precise" and "definite" were translated as "preciso" and "definito".

7.1.2 Analysis of each question's results

In the following sections, for easiness and to avoid repetitions, these terms will be used:

- "Slow Room": a room that generally features high values of Slowness, Inertia and/or Decay Time, in which the Response Envelopes of most peaks in the frequency response are not able to reach their actual steady state value with short bursts. An example of this type of rooms in this research is room CN.

- "Fast Room": a room that generally features low values of Slowness, Inertia and/or Decay Time, in which the Response Envelopes of most peaks in the frequency response are able to reach their actual steady state value even with short bursts. An example of this type of rooms in this research is room PRZ.

The naming of audio files is the same as the one presented in chapter 6, which is: TypeOfSound_Duration_Frequency_ConvolvedRoom. As an example, Bass_015_38_PRZ is a 0.15 second bass note at 38 Hz convolved with room PRZ's impulse response. If the last field is left empty, it means that the note was not convolved. Kick notes have duration 015 in their naming by default.

1. Kick_015_44_CN - Kick_015_38_PRZ

   (a) Which sound is the most precise?
   Testers who answered "First Sound": 5/30 = 16.67 %
   Testers who answered "Second Sound": 25/30 = 83.33 %

   (b) Which sound is the most resonant?
   Testers who answered "First Sound": 29/30 = 96.67 %
   Testers who answered "Second Sound": 1/30 = 3.33 %

   The question yielded significative results.

   Outcome of question 1: with a kick sound, 96.67 % of testers are able to correctly state that the kick convolved with the "slow" room's impulse response is more resonant than the kick convolved with the "fast" room's impulse response. 83.33 % of testers stated that the most precise kick sound is the one

convolved with the "fast" room's impulse response. It is clear that some confusion arises with the terms "precise" and "resonant", since the slight difference in the results means that these two terms are not complete opposites. In fact, a small number of people answered stating that the same sound was the most precise and the most resonant at the same time. Therefore, the second test will have to investigate further on the terms describing this phenomenon.

2. Bass_015_38_PRZ - Bass_015_44_CN

   (a) Which sound is the most precise?
   Testers who answered "First Sound": 27/30 = 90 %
   Testers who answered "Second Sound": 3/30 = 10 %

   (b) Which sound is the most resonant?
   Testers who answered "First Sound": 0/30 = 0 %
   Testers who answered "Second Sound": 30/30 = 100 %

   The question yielded significative results.

   Outcome of question 2: this confirms the results of question 1. With a short bass note, 100 % of testers are able to correctly state that the most resonant note is the one convolved with the "slowest" room's impulse response. The same ambiguity in the terms is also present for 10 % of testers.

3. Kick_015_44 - Kick_015_44_CN

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?
   Testers who answered "Worsened": 28/30 = 93.33 %
   Testers who answered "Remained equal": 2/30 = 6.67 %
   Testers who ansered "Improved": 0/30 = 0 %

   The question yielded significative results.

   Outcome of question 3: 93.33 % of testers have stated that the precision of the sound is worsened when comparing a synthetic dry kick drum hit with the same sound after the convolution with a "slow" room's impulse response. It is important to note that 0 % of testers have stated the opposite.

4. Kick_015_38 - Kick_015_38_PRZ

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?
   Testers who answered "Worsened": 22/30 = 73.33 %
   Testers who answered "Remained equal": 3/30 = 10 %
   Testers who ansered "Improved": 5/30 = 16.67 %

   The question yielded significative results.

   Outcome of question 4: 73.33 % of testers have stated that the precision of the sound is worsened when comparing a synthetic dry kick drum hit with the same sound after the convolution with a "fast" room's impulse response.

16.67 % of testers have stated that the precision improves after convolution. Comparing these results with those of question 3, there is a strong hint that the more "slow" is the room, the more testers are able to perceive a degradation in the precision of the sound (in this case, kick drum hits).

5. Kick_015_44 - Kick_015_44_CN - Kick_015_38 - Kick_015_38_PRZ

   (a) Is there a bigger precision degradation among sounds of the first couple (from the first to second sound) or the second couple (from the third to the fourth sound) ?

   Testers who answered "First Couple": 30/30 = 100 %

   Testers who answered "Second Couple": 0/30 = 0 %

   The question yielded significative results.

   Outcome of question 5: 100 % of testers stated that the "slow" room causes a bigger degradation in the perceived precision of the sound (in this case, kick drum hits).

6. Bass_015_44_CN - Bass_015_44

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

   Testers who answered "Worsened": 0/30 = 0 %

   Testers who answered "Remained equal": 5/30 = 16.67 %

   Testers who ansered "Improved": 25/30 = 83.33 %

   The question yielded significative results.

   Outcome of question 6: this confirms the results of question 3. 83.33 % of testers have stated that the precision of the non-convolved short bass note is higher than the precision of the same sound after the convolution with a "slow" room's impulse response. This percentage is slightly inferior than the one in question 3, but again, 0 % of testers stated the opposite.

7. Bass_015_38 - Bass_015_38_PRZ

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

   Testers who answered "Worsened": 7/30 = 23.33 %

   Testers who answered "Remained equal": 17/30 = 56.67 %

   Testers who ansered "Improved": 6/30 = 20 %

   The question yielded significative results.

   Outcome of question 7: More than 50 % of testers were not able to perceive a difference regarding the precision of a short bass note before and after the convolution with a "fast" room's impulse response. Only 23.33 % of testers were able to state that the convolved sound had a worse precision. This result further strenghtens the hypothesis that the fastest the room, the less testers are able to perceive its effects on precision of sounds.

8. Bass_015_44 - Bass_015_44_CN - Bass_015_38 - Bass_015_38_PRZ

   (a) Is there a bigger precision degradation among sounds of the first couple (from the first to second sound) or the second couple (from the third to the fourth sound) ?

   Testers who answered "First Couple": 30/30 = 100 %

   Testers who answered "Second Couple": 0/30 = 0 %

   The question yielded significative results.

   Outcome of question 8: this confirms the results of question 5, when tested with short bass notes: 100 % of testers stated that the "slow" room causes a bigger degradation in the perceived precision of the sound.

9. Bass_055_44 - Bass_055_44_CN

   (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

   Testers who answered "Worsened": 20/30 = 66.67 %

   Testers who answered "Remained equal": 5/30 = 16.67 %

   Testers who ansered "Improved": 5/30 = 16.67 %

   The question yielded significative results.

   Outcome of question 9: 66.67 % of testers stated that this sound (long bass notes) has worse precision after the convolution with a "slow" room's impulse response. 16.67 % of testers stated the opposite. Comparing these results with those of question 6 (which was the same question but with shorter bass notes), there's the hint that, the longest the note, the harder it is for testers to grasp the effect of room regarding the precision degradation.

10. Bass_055_38_PRZ - Bass_055_38

    (a) With respect to the first sound, the precision of the second sound is improved, worsened or remained equal?

    Testers who answered "Worsened": 5/30 = 16.67 %

    Testers who answered "Remained equal": 12/30 = 40 %

    Testers who ansered "Improved": 13/30 = 43.33 %

    The question yielded results which slightly contradict the behavior described previously, therefore more similar questions should be made to investigate the reason.

    Outcome of question 10: 43.33 % of testers stated that a long bass note loses precision after the convolution with a "fast" room's impulse response. 40 % did not notice any difference. Comparing this result with question 7, more people were able to answer correctly with longer bass notes. This behavior is opposite to the one found by questions 6 and 9, where less people could give the expected answer with longer notes. However, in questions 7 and 10 the second harmonic lies in room PRZ's big frequency response lack of support, while in question 6 and 9 the second harmonic lies on a peak in room CN's frequency response. While room PRZ is a "fast" room, behaving very well for short

sounds, if, according to the results obtained so far, it is true that when sounds are long enough the steady state curve could be more similar to the actual perception, this would mean that room PRZ could deliver a slightly different listening experience for longer sounds, more similar to the one depicted by the Steady State Curve. While the precision probably does not change, testers may hear a difference (wether that lies in timbre, precision, or other parameters) and answer differently. In this case, the effect of the duration of the note on the perceived precision would depend on the specific frequency response of the room and on the sensibility of the person that it listening. This phenomenon should be object of further investigation.

What is ultimately important, though, is that, once again, the "faster" room yelds less perceptible precision degradation with respect to the "slowest" room.

11. Bass_055_44 - Bass_055_44_CN - Bass_055_38 - Bass_055_38_PRZ

    (a) Is there a bigger precision degradation among sounds of the first couple (from the first to second sound) or the second couple (from the third to the fourth sound) ?

    Testers who answered "First Couple": 22/30 = 73.33 %

    Testers who answered "Second Couple": 8/30 = 26.67 %

    The question yielded significative results.

    Outcome of question 11: 73.33 % of testers stated that, on long bass notes, the "slow" room introduces a bigger loss of perceived precision. Comparing this result with questions 5 and 8, there is the hint that this effect can be less perceptible with longer notes.

12. Bass_003_72 - Bass_003_72_PRZ (on a valley) - Bass_003_72_DrmREG (on a peak)

    (a) Is the pitch always the same?

    (b) With how much confidence would you say this? (from 1 to 10)

    Testers who answered "Yes": 15/30 = 50 % (Avg. Confidence = 7.733/10)

    Testers who answered "No": 15/30 = 50 % (Avg. Confidence = 6.467/10)

    The question yielded significative results.

    Outcome of question 12: 50 % of testers do not perceive the notes as having the same pitch (some testers stated that the pitch was descending from the first to the third sound). This means that pitch perception is difficult for very short notes, but it is also possible that strong resonances can modify the perception of the pitch of really short (30 msec) bass notes, because some listeners stated that the pitch of the sounds were descending from the first to the third.

13. Bass_015_72 - Bass_015_72_PRZ (on a valley) - Bass_015_72_DrmREG (on a peak)

    (a) Is the pitch always the same?

(b) With how much confidence would you say this? (from 1 to 10)
Testers who answered "Yes": 25/30 = 83.33 % (Avg. Confidence = 8.32/10)
Testers who answered "No": 5/30 = 16.67 % (Avg. Confidence = 8/10)

The question yielded significative results.

Outcome of question 13: Comparing these results with question 12, more people are able to correctly state that the pitch is the same when the note is slightly longer (150 msec).

14. Bass_055_72 - Bass_055_72_PRZ (on a valley) - Bass_055_72_DrmREG (on a peak)

(a) Is the pitch always the same?

(b) With how much confidence would you say this? (from 1 to 10)
Testers who answered "Yes": 27/30 = 90 % (Avg. Confidence = 8.148/10)
Testers who answered "No": 3/30 = 10 % (Avg. Confidence = 6.633/10)

The question yielded significative results.

Outcome of question 14: Comparing these results with question 12 and 13, even more people are able to correctly state that the pitch is the same when the note is longer (550 msec). Also, testers who stated the opposite, did so with lower confidence than the previous questions.

Questions 12, 13 and 14 give a strong hint regarding the fact that the presence of a room mode may alter the perception of the pitch for really short sounds.

15. Kick_015_98_DrmB - Kick_015_49_DrmB - Kick_15_196_DrmB

(a) Which sound is the most resonant?
Testers who answered "First sound": 1/30 = 3.33 %
Testers who answered "Second sound": 27/30 = 90 %
Testers who answered "Third sound": 2/30 = 6.67 %

(b) Which sound is more precise/defined?
Testers who answered "First sound": 11/30 = 36.67 %
Testers who answered "Second sound": 3/30 = 10 %
Testers who answered "Third sound": 16/30 = 53.33 %

The question yielded significative results.

Outcome of question 15: 90 % of testers stated that the most resonant sound is the one whose fundamental frequency is centered on the highest and lowest room mode, as expected. 53.33 % of testers have stated that the most precise/defined of the three is the one whose fundamental is one octave above the lowest room mode, while 36.67 % stated that it is the one whose fundamental is one octave below. Probably, Equal Loudness Curves played a role in question 15(b), giving the impression that the third sound was louder than the first one. This could be investigated further. What is important, though, is that the Kick hit with its fundamental on the lowest room mode was perceived as being the most resonant and less precise / defined.

16. Bass_015_99_PRZ - Bass_015_111_PRZ

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)
    Testers who answered "Yes": 24/30 = 80 % (Avg. Confidence = 7.75/10)
    Testers who answered "No, first sound is quieter": 6/30 = 20 % (Avg. Confidence = 7.333/10)
    Testers who answered "No, second sound is quieter": 0/30 = 0 %

    The question yielded significative results.

    Outcome of question 16: 80 % of testers perceived the same volume, meaning that on short notes the attack plays an important role in volume perception (the steady state of the first sound was really close in amplitude to the overshoot level of the second sound). 0 % of testers stated that the note on the valley was quieter.

17. Bass_025_99_PRZ_NoOvershoot - Bass_025_111_PRZ_NoOvershoot

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)
    Testers who answered "Yes": 18/30 = 60 % (Avg. Confidence = 7.611/10)
    Testers who answered "No, first sound is quieter": 1/30 = 3.33 % (Avg. Confidence = 5/10)
    Testers who answered "No, second sound is quieter": 11/30 = 36.67 % (Avg. Confidence = 7.636/10)

    The question yielded significative results.

    Outcome of question 17: by manually removing the initial and final part of both sounds, therefore removing the overshoot portion, less testers than question 16 perceived the same volume, while 36.67 % of them started to hear as quieter the one with lower steady state value, hinting that the overshoot portion could play an important role when evaluating the perceived level of short notes.

18. Bass_015_111_PRZ - Bass_025_111_PRZ_NoOvershoot

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)
    Testers who answered "Yes": 6/30 = 20 % (Avg. Confidence = 6.5/10)
    Testers who answered "No, first sound is quieter": 2/30 = 6.67 % (Avg. Confidence = 6.5/10)
    Testers who answered "No, second sound is quieter": 22/30 = 73.33 % (Avg. Confidence = 7.995/10)

    The question yielded significative results.

    Outcome of question 18: by direct comparison between the second sound of question 16 and the second sound of question 17 (which is the same bass note

with and without Overshoots) 73.33 % of testers perceived as quieter the one without overshoots. This confirms that the Overshoot portion is important when evaluating levels of short notes.

19. Bass_015_246_SGR - Bass_015_280_SGR

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)
    Testers who answered "Yes": 13/30 = 43.33 % (Avg. Confidence = 7.463/10)
    Testers who answered "No, first sound is quieter": 1/30 = 3.33 % (Avg. Confidence = 8/10)
    Testers who answered "No, second sound is quieter": 16/30 = 53.33 % (Avg. Confidence = 7/10)

    The question yielded non significative results.

    Outcome of question 19: 53.33 % of testers perceived the second sound (on the valley) to have a lower volume, while 43.33 % perceived them as having the same volume.

20. Bass_055_246_SGR - Bass_055_280_SGR

    (a) Do these sounds have the same volume? If not, which one is the quietest?

    (b) With how much confidence would you say this? (from 1 to 10)
    Testers who answered "Yes": 12/30 = 40 % (Avg. Confidence = 7.583/10)
    Testers who answered "No, first sound is quieter": 3/30 = 10 % (Avg. Confidence = 6/10)
    Testers who answered "No, second sound is quieter": 15/30 = 50 % (Avg. Confidence = 7.467/10)

    The question yielded non significative results.

    Outcome of question 20: testing with longer bass notes with respect to question 19, results are almost the same. It was expected that more people would be able to perceive the second sound as being quieter with respect to the previous question. It appears that the note duration does not have an immediate effect on the perceived volume of notes. This would go against the hypothesis that gives high importance to the presence of overshoots when evaluating the loudness of short sounds. However, both question 19 and 20 were performed in the higher part of the chosen frequency range. This could introduce other effects, therefore these questions will be repeated in the second test at lower frequencies.

21. Bass_015_111_PRZ - Bass_025_111_PRZ_- Bass_055_111_PRZ

    (a) Do these sounds have the same volume? If not, which one is the loudest?

    (b) With how much confidence would you say this? (from 1 to 10)
    Testers who answered "Yes": 24/30 = 80 % (Avg. Confidence = 7.5/10)
    Testers who answered "No, first sound is the loudest": 0/30 = 0 %

Testers who answered "No, second sound is the loudest": 2/30 = 6.67 % (Avg. Confidence = 7/10)

Testers who answered "No, third sound is the loudest": 4/30 = 13.33 % (Avg. Confidence = 8.25/10)

The question yielded significative results.

Outcome of question 21: the three sounds were all equal besides their duration, and had their fundamental frequency on a valley in the frequency response of a "fast" room. 80 % of testers perceived them as having the same volume, while 6 testers perceived them as having different loudness.

22. Bass_015_44_CN - Bass_025_44_CN_- Bass_055_44_CN

    (a) Do these sounds have the same volume? If not, which one is the loudest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Testers who answered "Yes": 18/30 = 60 % (Avg. Confidence = 7.722/10)

    Testers who answered "No, first sound is the loudest": 1/30 = 3.33 % (Avg. Confidence = 9/10)

    Testers who answered "No, second sound is the loudest": 4/30 = 13.33 % (Avg. Confidence = 5/10)

    Testers who answered "No, third sound is the loudest": 7/30 = 23.33 % (Avg. Confidence = 7.286/10)

    The question yielded results that are not completely significative, but hint at a particular phenomenon.

    Outcome of question 22: with the same scenario as in question 21, but with sounds having their fundamental frequency on a peak in the frequency response of a "slow" room, less testers perceived them as having the same volume. A small percentage of testers perceived the longest one as being the loudest. Combining the results of questions 21 and 22, there is a slight hint at the fact that, the longer the note, the higher is the perceived volume (when the note is the same and the presence of overshoots is not altered). However, this would need to be developed through further research as results of questions 21 and 22 are too inconsistent to state anything with certainty.

23. Bass_055_44 - Bass_055_44_CN

    (a) Do these sounds have the same volume? If not, which one is the loudest?

    (b) With how much confidence would you say this? (from 1 to 10)

    Normal listening level: Testers who answered "Yes": 6/30 = 20 % (Avg. Confidence = 6.667/10)

    Testers who answered "No, first sound is the loudest": 19/30 = 63.33 % (Avg. Confidence = 7.105/10)

    Testers who answered "No, second sound is the loudest": 5/30 = 16.67 % (Avg. Confidence = 7/10)

    Louder listening level:

    Testers who answered "Yes": 13/30 = 40 %

Testers who answered "No, first sound is the loudest": 9/30 = 30 %

Testers who answered "No, second sound is the loudest": 8/30 = 26.67 %

The question yielded significative results, but the change in volume was outside the scope of the test. This phenomenon should be investigated further with other psychoacoustic tests.

Outcome of question 23: at normal listening level, 63.33 % of testers stated that the loudest sound was the non-convolved one. However, at higher listening level, the results shifted towards the convolved sound. In fact, the percentage of testers who perceived the first one as being louder dropped, while testers who perceived the convolved one as louder slightly increased and testers who perceived them as having the same volume increased. This is a hint at the fact that the volume perception is strongly influenced both by the listening level and the "precision" of the sound. The author thinks that at lower volume the non-convolved sound was perceived as being louder because of its attack being non-compromised by the room mode of the slow room. This would result in a higher initial volume. However, for louder listening levels, a phenomenon similar to the "energization" of sounds suggested in M. Ferroni's thesis work could happen, leading to the impression that the sound with fundamental frequency at the slowest room mode is actually louder after the convolution. Of course, both files were normalized with respect to their peak as explained in chapter 6 so that the perceived difference would be only caused by the convolution phenomenon rather than an actual level difference in the files.

7.1.3 Conclusions of the First Test

- Some testers have asked questions about the meaning of the words "precision" and "definition" related to sound. Further work on the vocabulary is needed in the second test.

- Kick notes have a complex and wide spectrum, while Bass notes have an harmonic spectrum. Because of the difference in their spectral content, different sounds can highlight or hide different perceptual behaviors regarding all curves defined so far, yelding slightly different results.

- Testers are, for the majority, able to perceive distinctly the precision degradation after the convolution of a sound with the impulse response of a "slow" room.

- Testers are not always able to perceive the precision degradation after the convolution of a sound with the impulse response of a "fast" room.

- Testers are able to address which room degrades the most the perceived precision of a sound, if using a direct comparison and two rooms with different temporal characteristics. In particular, testers were always correct with kick and short bass notes, while a slightly lower percentage of correct answers was given with long bass notes.

- Conflicting results arise when evaluating the perceived precision on the same note and room when varying the note duration. The specific circumstances of the questions and results suggest that the listening experience for longer notes is slightly modified according to the Steady State Curve, therefore depending on the specific room's characteristics. Further research should be done to understand this phenomenon.

- Room modes heavily influence the perception of the pitch of really short tones centered at the frequency of the modes.

- Given a room mode, and three notes which are centered at the same frequency, one octave below and one octave above, the one which is perceived as being the most resonant is the first one, while the one which is perceived as being the most precise is the third one.

- When comparing two short sounds of which one has its fundamental on a room mode, and the other has its fundamental on a valley of the room's frequency response (therefore, showing overshoot behavior), if the overshoot value is similar to the steady state value of the first sound, most testers perceive both sounds as having the same volume. Therefore, the classic frequency response curve appears to not be a good reference regarding the level perception of short sounds.

- In the same scenario, If the initial and final parts (containing the overshoots) of the second sounds are removed, more testers are able to state that the sound on the valley is quieter than the one on the room mode, even though the majority still perceives them as being equally loud.

- When comparing directly a sound whose fundamental frequency is on a frequency response's valley, with the same sound after removing overshoots, most testers can correctly state that the second one is quieter. This means that the overshoots play a part when assessing the loudness of short notes.

- There is a hint on the fact that, comparing the same note with different durations, the louder is perceived by a small number of testers. This is true for notes whose fundamental frequency is on valleys (leaving their overshoot behavior untouched) as well as on peaks. However, this behavior could be strongly influenced by the position of overtones in the frequency response, especially for sounds with a complex spectrum, such as kick drums, potentially augmenting the overall volume. This phenomenon should be studied in more detail before drawing conclusions.

Regarding the perceived precision and the effect of overshoots, and summarizing the previous results, it appears that:

- When a short sound is played in a "fast" room, the Overshoot Response appears to be more closely related to the loudness psychoacoustic perception than the classic frequency response curve is, and room modes affect very slightly the perceived precision of the sound.

- When a short sound is played in a "slow" room, the Overshoot Response appears to be more closely related to the loudness psychoacoustic perception than the classic frequency response curve is, and room modes affect heavily the perceived precision of the sound.

- The effect of room modes on precision and definition loss seems to depend mainly on the temporal characteristics of the room (Inertia, Slowness and Decay Time values).

- While it has not been a direct consequence of one question's results, the high percentage of testers who perceived the removal of the overshoots seem to suggest that the longer the note, the more the perception should follow the Steady State Curve and Frequency Response. Questions 19 and 20 seem to suggest otherwise, but they were performed at higher frequencies than the rest of the questions, so they will be repeated in the second test at lower frequencies, where all other behaviors and questions have been evaluated.

7.2 Second Test results

After developing Room Slowness, Inertia and Decay Time calculations, the second psychoacoustic test was performed and results were analyzed.

7.2.1 Tester's statistics

Again, all testers were musicians or expert listeners. Eighteen testers took part in both tests, while the others were not available for both.

Number of testers: 30
Number of bass players, drummers or producers among testers: 16
Average age of testers: 25,63 years
Younger tester: 19 years
Older tester: 38 years
Male testers: 29
Female testers: 1
People who experienced listening fatigue at the end of the test: 0
People who were annoyed / felt uncomfortable because of the many questions: 0

7.2.2 Analysis of each question's results

1. Kick_015_32_PRZ - Kick_015_38_PRZ

   (a) Which sound is the most precise?
   Testers who answered "First Sound": 19/30 = 63.3 %
   Testers who answered "Second Sound": 11/30 = 36.7 %

   (b) Which sound is the most resonant?
   Testers who answered "First Sound": 23/30 = 76.7 %
   Testers who answered "Second Sound": 7/30 = 23.3 %

   The question yielded significative results.

   Outcome of question 1: the two kick drum sounds were really close to each other and the question aimed at verifying if the difference in Slowness and Inertia of a single frequency (the fundamental) was audible. It appears that it is not, since the note with higher Slowness is the first one. However, this is expected since the spectral complexity of a kick hit may hide the behavior of one specific frequency.

   Also, the fact that the first sound was, for some testers, both "precise" and "resonant" at the same time confirms the need to tune the vocabulary and reformulate the test.

2. Bass_055_32_PRZ - Bass_055_38_PRZ

   (a) Which sound is the most precise?
   Testers who answered "First Sound": 5/30 = 16.67 %
   Testers who answered "Second Sound": 25/30 = 83.3 %

(b) Which sound is the most resonant?
Testers who answered "First Sound": 25/30 = 83.3 %
Testers who answered "Second Sound": 5/30 = 16.67 %

The question yielded significative results. Outcome of question 2: in the same scenario as question 1, but testing with long bass notes, 83.3 % of testers were able to point out the most precise note, which is the one tuned at the fundamental frequency with lowest Slowness and Inertia values. The reason could be that the spectrum of the bass is harmonic. Being simpler than the spectrum of a kick, it probably generates less masking than kick hits.

3. Kick_015_58_PRZ - Kick_015_32_PRZ

(a) Which sound is the most precise?
Testers who answered "First Sound": 17/30 = 56.67 %
Testers who answered "Second Sound": 13/30 = 43.3 %

(b) Which sound is the most resonant?
Testers who answered "First Sound": 11/30 = 36.7 %
Testers who answered "Second Sound": 19/30 = 63.3 %

Outcome of question 3:

The question yielded non significative results.

More than half testers correctly stated that the most precise sound was the one whose fundamental lied in a valley in the frequency response (whereas the other one was tuned on a peak). However, the percentage in this case is really close to 50 %, leaving some doubt about the validity of this result. However, from both questions 1 and 3, it is clear that evaluating the precision of different kick sounds in the same room provides results that vary slightly, while, from test 1, evaluating the precision of different rooms using kick sounds is perceived a lot more clearly by testers. This strenghtens the idea that the precision degradation is not modified by only a single frequency, but the most immediate difference is given by the global room's conditions.

4. Kick_015_52_SNT - Kick_015_44_CN - Kick_015_38_PRZ - Kick_015_38_SGR

(a) Rank the four sounds from the most precise/defined to the least precise/defined.

In the following, D is room SGR, C is room PRZ, B is room CN, A is room SNT.
Testers who answered "ADCB": 1/30 = 3.33 %
Testers who answered "ADBC": 1/30 = 3.33 %
Testers who answered "BCDA": 1/30 = 3.33 %
Testers who answered "CDBA": 5/30 = 16.7 %
Testers who answered "CDAB": 1/30 = 3.33 %
Testers who answered "CBDA": 2/30 = 6.66 %
Testers who answered "DCBA": 15/30 = 50 %
Testers who answered "DCAB": 4/30 = 13.3 %

The question yielded significative results.

Outcome of question 4: 50 % of testers ranked the rooms in the order SGR - PRZ - CN - SNT. This ranking is the same as the one found in chapter 5.12 using the Inertia parameter. 16.67 % of testers ranked the rooms in the order PRZ - SGR - CN - SNT, which is the same ranking as the one obtained with both Room Slowness and Decay Time parameters, and 13.3 % ranked the rooms in order SGR - PRZ - SNT - CN. In this case, it looks like Inertia is the room parameter that can describe the perceived loss of precision after the convolution with a room impulse response. To a lesser degree, Slowness and Decay Time also give similar results. However, this is surprising, since Room Inertia is not, by definition, well suited to describe non-sustained sounds (as introduced in chapter 5.10.3). Question 12, however, will introduce different results.

Note that some testers said that room D (SNT) was so different from the others that they could not decide how to classify it.

5. Kick_015_38_SGR - Kick_015_32_SGR

   (a) Do both sounds have the same volume? if not, which one is the quietest?
   Testers who answered "Yes": 20/30 = 66.67 %
   Testers who answered "No, first sound is the quietest": 3/30 = 10 %
   Testers who answered "No, second sound is the quietest": 7/30 = 23.33 %

   The question yielded significative results.

   Outcome of question 5: The reader is reminded that room SGR has a valley at 32 Hz and a peak at 38 Hz, with equal overshoot level and different steady state level. 66.67 % of testers perceived the same volume, while 23.33 % perceived the one with fundamental frequency on the valley as being the quietest.

6. Bass_015_38_SGR - Bass_015_32_SGR

   (a) Do both sounds have the same volume? if not, which one is the quietest?
   Testers who answered "Yes": 23/30 = 76.7 %
   Testers who answered "No, first sound is the quietest": 0/30 = 0 %
   Testers who answered "No, second sound is the quietest": 7/30 = 23.33 %

   The question yielded significative results.

   Outcome of question 6: repeating question 5 with short bass notes, the results are almost the same. No one stated that the quietest one was the one with fundamental on the room's frequency response's peak. The majority of testers perceived them as having the same volume.

7. Bass_055_32_SGR - Bass_055_38_SGR

   (a) Do both sounds have the same volume? if not, which one is the quietest?
   Testers who answered "Yes": 14/30 = 46.67 %
   Testers who answered "No, first sound is the quietest": 16/30 = 53.3 %

Testers who answered "No, second sound is the quietest": 0/30 = 0 %

The question yielded significative results.

Outcome of question 7:

Questions 6 and 7 were the remake of questions 19 and 20 of test one.

Repeating question 5 with longer bass notes, more than 50 % of testers were now able to state that the quietest one was the one with fundamental frequency in the valley. The percentage of testers who heard them with the same volume dropped from question 6 from 76.7 % to 46.67 % and no one stated that the quietest one was the one with fundamental on the room's frequency response's peak.

While the number of testers who heard the sounds with same volume and those who pointed out the quietest one were really close, the change with respect to the previous question is clear and this is a hint on the fact that, the longer the note, the more the Steady State Curve represents the perception of sustained notes.

8. Bass_015_38_SGR_noOvershoot - Bass_015_32_SGR_noOvershoot

   (a) Do both sounds have the same volume? if not, which one is the quietest?
   Testers who answered "Yes": 17/30 = 56.7 %
   Testers who answered "No, first sound is the quietest": 3/30 = 10 %
   Testers who answered "No, second sound is the quietest": 10/30 = 33 %

   The question yielded significative results.

   Outcome of question 8: Compared to question 6, the removal of overshoot from both sounds decreases the percentage of testers who perceive them at the same volume, and increases the percentage of testers who perceive them differently. A small percentage of testers (10 %), however, perceived as louder the one with higher steady state value.

9. Bass_015_32_SGR - Bass_015_32_SGR_noOvershoot

   (a) Do both sounds have the same volume? if not, which one is the quietest?
   Testers who answered "Yes": 15/30 = 50 %
   Testers who answered "No, first sound is the quietest": 0/30 = 0 %
   Testers who answered "No, second sound is the quietest": 15/30 = 50 %

   The question yielded significative results.

   Outcome of question 9: Half of the testers perceived the sounds at the same level, while the other half correctly pointed out the quieter one. The fact that no one perceived as quieter the sound with overshoot confirms the result of question 18 of test 1, regarding the importance of the overshoot presence in the level perception of low frequency short sounds.

10. PureTone_055_32_SGR - PureTone_055_38_SGR

(a) Do both sounds have the same volume? if not, which one is the quietest?
Testers who answered "Yes": 2/30 = 6.6 %
Testers who answered "No, first sound is the quietest": 24/30 = 80 %
Testers who answered "No, second sound is the quietest": 4/30 = 13.3 %

The question yielded significative results.

Outcome of question 10: 80 % of testers were able to point out the quieter sound between two (quite long) pure tones convolved with the room impulse response at the same frequencies of the previous questions. The fact that there is no masking caused by other spectral components has made quite clear which sound was the quietest. Furthermore, testers clearly perceived two "peaks" in the second sound, which correspond to the overshoots, and were asked to evaluate the sound level in its entirety.

11. Bass_015_44 - Bass_015_44_CN

Bass_015_36 - Bass_015_36_SGR

Bass_015_38 - Bass_015_38_PRZ

Bass_015_52 - Bass_015_52_SNT

(a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.
Average rating of room PRZ: 84.4/100
Average rating of room SGR: 72.57/100
Average rating of room CN: 44.37/100
Average rating of room SNT: 20.07/100

The question yielded significative results.

Outcome of question 11: with short bass notes, the preferred ranking for rooms with respect to the precision and definition of the sound was PRZ - SGR - CN - SNT, corresponding to the ranking provided by Room Slowness and Decay Time.

12. Bass_055_38 - Bass_055_38_PRZ

Bass_055_44 - Bass_055_44_CN

Bass_055_52 - Bass_055_52_SNT

Bass_055_36 - Bass_055_36_SGR

(a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.
Average rating of room PRZ: 74.993/100
Average rating of room SGR: 75.233/100
Average rating of room CN: 55.52/100
Average rating of room SNT: 33.5/100

The question yielded significative results.

Outcome of question 12: with long bass notes, the preferred ranking for rooms with respect to the precision and definition of the sound was SGR- PRZ - CN - SNT, corresponding to the ranking provided by Room Inertia (even though the very small difference between the average ratings of rooms SGR and PRZ is not enough to give a certain answer).

13. Kick_015_36 - Kick_015_36_SGR

    Kick_015_52 - Kick_015_52_SNT

    Kick_015_38- Kick_015_38_PRZ

    Kick_015_44 - Kick_015_44_CN

    (a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.

    Average rating of room PRZ: 69.95/100

    Average rating of room SGR: 53.5/100

    Average rating of room CN: 29.53/100

    Average rating of room SNT: 23.37/100

    The question yielded significative results.

    Outcome of question 13: with kick drum notes, the preferred ranking for rooms with respect to the precision and definition of the sound was PRZ - SGR - CN - SNT, corresponding to the ranking provided by Room Slowness and Decay Time. This is in contrast with the results of question 4, which yelded as result the ranking SGR - PRZ - CN - SNT, the same as the one generated by Room Inertia. Probably, the different presentation of the questions led testers to give different answers. Question 4 asked to rank them with respect to each other, while question 13 also allowed to hear the comparison between the convolved sound and an anchor. Further studies should be made in order to understand why this happened.

14. Excerpt_44 - Excerpt_44_CN

    Excerpt_38 - Excerpt_38_PRZ

    Excerpt_52 - Excerpt_52_SNT

    Excerpt_36 - Excerpt_36_SGR

    (a) For each couple of sounds, take as reference the first one (as if its precision / definition was the maximum of the scale) and draw a vertical line on the scale on the point that represents the precision of the second sound with respect to the first one.

    Average rating of room PRZ: 73.967/100

    Average rating of room SGR: 66.8/100

    Average rating of room CN: 37.87/100

    Average rating of room SNT: 22.7/100

The question yielded significative results.

Outcome of question 14: with musical excerpts, the preferred ranking for rooms with respect to the precision and definition of the sound was PRZ - SGR - CN - SNT, corresponding to the ranking provided by Room Slowness and Decay Time.

15. Excerpt_44_CN - Excerpt_38_PRZ - Excerpt_52_SNT - Excerpt_36_SGR

    (a) For each sound, draw a vertical line on the scale on the point that represents your subjective preference of that sounds' listening quality.
    Average rating of room PRZ: 70.067/100
    Average rating of room SGR: 78.1/100
    Average rating of room CN: 29.9/100
    Average rating of room SNT: 19.467/100

The question yielded significative results.

Outcome of question 15: with musical excerpts, the preferred ranking for rooms with respect to the overall preference of the sound was SGR - PRZ - CN - SNT. Besides temporal parameters, since this question regarded the sound in all its aspects (ingoring the actual notes being played), the author's hypothesis is that if rooms are very "slow", the preference is of course low as their impact on the sound precision and definition is intense. However, when rooms are "fast", the impact of the convolution is not so annoying, and other aspects can be more important. In fact, the reader is reminded that the frequency response of room PRZ has a huge valley, whereas the frequency response of room SGR is a lot flatter. That could be the reason while, overall, room SGR was preferred to room PRZ. Some testers expressed that room PRZ had a poor bass response.

16. Bass_015_44 - Bass_015_44_- CN

    (a) For each sound, describe it with a number of adjectives of your choice from the list, or write one or two new words that you think describes it better.
    "Definito, Cupo, Confuso, Asciutto, Preciso, Risonante, Riverberato, Pulito, Distorto, Smorzato, Gonfio, Sporco, Rimbombante"

The question yielded significative results.

Outcome of question 16:

Sound 1 (non convolved):

Number of testers that described it as "Definito": 22

Number of testers that described it as "Preciso": 17

Number of testers that described it as "Asciutto": 16

Number of testers that described it as "Pulito": 14

Number of testers that described it as "Smorzato": 9

Number of testers that described it as "Distorto": 2

Number of testers that described it as "Sporco": 2

Number of testers that described it as "Gonfio": 1

New entries: "Secco" (1 tester), "Stoppato" (1 tester), "Netto" (1 tester), "Troncato" (1 tester)

Regarding the word "Distorto" (distorted), both testers mentioned they were referring to the timbre of the sound, which was a bass played with a plectrum. While the sound did not feature any distortion (neither digital nor simulated by a virtual instrument), it is possible that they were referring to the attack of the plectrum on the strings.

Sound 2 (convolved with a "slow" room):

Number of testers that described it as "Riverberato": 26

Number of testers that described it as "Rimbombante": 17

Number of testers that described it as "Gonfio": 9

Number of testers that described it as "Confuso": 8

Number of testers that described it as "Risonante": 8

Number of testers that described it as "Sporco": 5

Number of testers that described it as "Smorzato": 3

Number of testers that described it as "Definito": 3

Number of testers that described it as "Preciso": 3

Number of testers that described it as "Distorto": 2

Number of testers that described it as "Cupo": 2

Number of testers that described it as "Pulito": 1

New entries: no new entries

17. Bass_015_38_PRZ - Bass_015_44_- CN

    (a) For each sound, describe it with a number of adjectives of your choice from the list, or write one or two new words that you think describes it better.

    "Definito, Cupo, Confuso, Asciutto, Preciso, Risonante, Riverberato, Pulito, Distorto, Smorzato, Gonfio, Sporco, Rimbombante"

The question yielded significative results.

Outcome of question 17:

Sound 1 (convolved with a "fast" room):

Number of testers that described it as "Definito": 26

Number of testers that described it as "Preciso": 19

Number of testers that described it as "Pulito": 16

Number of testers that described it as "Asciutto": 7

Number of testers that described it as "Risonante": 3

Number of testers that described it as "Gonfio": 2

Number of testers that described it as "Cupo": 1

Number of testers that described it as "Smorzato": 1

Number of testers that described it as "Riverberato": 1

New entries: "Medioso" (1 tester), "Filtrato" (1 tester)

Sound 2 (convolved with a "slow" room):

Number of testers that described it as "Rimbombante": 17

Number of testers that described it as "Cupo": 16

Number of testers that described it as "Gonfio": 11

Number of testers that described it as "Smorzato": 9

Number of testers that described it as "Riverberato": 8

Number of testers that described it as "Risonante": 8

Number of testers that described it as "Confuso": 7

Number of testers that described it as "Sporco": 3

Number of testers that described it as "Distorto": 2

Number of testers that described it as "Preciso": 2

Number of testers that described it as "Pulito": 1

New entries: "Lontano" (1 tester), "Ovattato" (2 testers)

7.2.3 ANOVA Analysis

A two-way ANOVA analysis was run with the data of questions 11, 12, 13, 14 of test 2. The two variables under test were "Type of sound" (kick drums hits, short bass notes, long bass notes, musical excerpts) and "Temporal behavior of the room" (really fast, fast, slow, really slow). The temporal behavior of the room referred mainly to the ranking provided by Room Slowness and Decay Time. It is important to remind that, since impulse responses of real rooms were used, the variables regarding the room were not controllable. Therefore, a phenomenologic analysis was made, and room PRZ was chosen as "really fast" (lower Slowness / Decay Time values), room SGR was chosen as "fast" (lower Inertia value), room CN was chosen as "slow", room SNT was chosen as "really slow" (higher Slowness / Decay Time values). Of course this is a semplification aimed at evaluating the global temporal behavior of the room, but further research can be aimed at studying the specific dependance of these results on singular parameters of the room, possibly using synthetic models in order to control only one variable at a time.

The two-way ANOVA analysis was performed through a Matlab™ script developed by the author. A confidence level of 5% was used. The result is as follows:

- F First Factor (Type of Sound):5.807430e+00

  p-value First Factor: 6.672117e-04

  Confidence Level: 5.000000e-02

Null Hypothesis can be rejected because p-value < confidence level.

This means that the Type of Sound is significant to the results.

- F Second Factor (Room Temporal Behavior):2.834379e+02

  p-value Second Factor: 0

  Confidence Level: 5.000000e-02

  Null Hypothesis can be rejected because p-value < confidence level.

  This means that the Room Temporal Behavior is significant to the results.

- F Both Factors:-1.071713e+02

  p-value Both Factors: 1

  Confidence Level: 5.000000e-02

  No significant evidence against Null Hypothesis because p-value > confidence level.

  This means that the interaction between Type of Sound and Room Slowness is not significant to the results.

These results indicate that:

- The type of sound is significative for the rating that testers give to each room, meaning that different sounds are rated differently in the same room.
- The temporal behavior of the room (levels of Room Slowness and Decay Time) is significative for the rating that testers give to each room, meaning that the same sound is rated differently in different rooms.
- The interaction between Type of Sound and Room Slowness is not significant to the results. This means that one value of a variable has always the same trend with respect to the other one. As an example, a slower room tends to generate lower precision ratings for all levels of the other variable "Type of sound". If this wasn't true and interaction between variables was present, it would mean that, as an example, slower rooms would give lower ratings for certain types of sounds and higher ratings for others.

7.2.4 Room Preferences

It appears, as expected, that the preferred rooms are those with a "fast" behavior (in which the response envelopes reach their steady state even on frequency response peaks and for short notes) rather than "slow" rooms. The temporal behavior of rooms both in their rise time (Room Slowness) and in their final part (Room Inertia and Room Decay) all impact the perceived precision and definition. Slightly contrasting results arise from different questions regarding which, if any, of these variable is more important than the others in the perceptual domain. Therefore, further studies should be made to inspect this behavior. However, it appears that the new room metrics Room Slowness and Room Inertia can be used alongside the Decay Time as another tool of inspection, one which is more representative of the specific behavior of each single frequency if compared with the Decay Time.

As far as general preference goes, the perceived precision and definition plays a major role. However, there is a hint on the fact that other variables could be relevant, such as the frequency response curve: if the sounds are long enough (which is the case for musical excerpts) the author suggests that the classic frequency response curve is more representative than the flatter overshoot response (which is, instead, more significative for very short sounds), leaving all imperfections of the frequency response to be audible to the listener, potentially changing the preference of that specific listening experience.

7.2.5 Vocabulary Analysis

Regarding the words used in both psychoacoustic tests and the final analysis in test 2, the reader is reminded that the initial pool of words was developed in M. Ferroni's thesis work with specific psychoacoustic tests. Since some ambiguity arised in test 1, some questions in test 2 aimed at further confirming this pool of words. As a matter of fact, no new significative entries were introduced: of 8 new entries, only one was entered two times. Regarding the sounds that were defined as "precise" (basically, the ones which were not convolved or convolved with "fast" rooms) showed that the most significative term in both questions was "definito" (italian for "well-defined", "sharp", "clear"). The word "preciso" (precise) was second to that in both cases. Other significative words in both questions were "asciutto" (dry) and "pulito" (clean). However, these terms are more descriptive of the timbre of the sound and the presence of effects such as reverb. Therefore, the author suggest against using these words and using "definito" and "preciso" as initially suggested.

Regarding the sounds that were convolved with slow rooms (therefore, which showed a loss of precision and definition), the most significative term in one question was "riverberato" (reverberated), followed by "rimbombante" (booming). In the second questions, it was "rimbombante" followed by "cupo" (dark), "gonfio" (which means inflated, but is used in italian as a synonym of "cupo"). "reverberated" was, in the second question, lower in the ranking. The word "risonante" (resonant) was initially used in the tests, but was not chosen as much as other adjectives. In italian, "risonante" and "rimbombante" are almost synonyms. However, "risonante" is more technical and could be more understandable by experts and musicians, whereas "rimbombante" is more used in everyday's language. Therefore, it is suggested that the words "rimbombante" (booming) is used alongside "risonante" (resonant) to describe the loss in perceived precision and definition caused by the convolution with "slow" rooms' impulse responses. Again, the author suggests to ignore the word "riverberato" (reverberated) as it is more descriptive of the timbre of the sound, and the presence of a reverb. Also, this word ranked at the first position for one question, while ranking at the fifth for the second question. "Rimbombante", however, ranked second and first, respectively.

7.2.6 Conclusions of the Second Test

- the complexity of the spectrum of a sound can mask the behavior of a limited number of low-frequency problems, such as peaks in the frequency response. Sounds with simpler spectrums seem to mask these problems in a less pronounced way.

- Again, short sounds with fundamental frequencies on peaks and valleys with a similar Overshoot level were perceived by most testers as having the same volume. With longer notes, more testers could point out that the second was quieter, hinting at the fact that the room frequency response could be perceptually more relevant for long sounds, while the overshoot response could be more appropriate for short sounds.

- In the same scenario as the previous point, if the initial and final parts were removed leaving their steady state exposed, more testers would correctly point out that the second sound was quieter. This further confirms the importance of Overshoots in the perception of low frequency short sounds.

- Comparing one short sound with overshoots, and the same sound after the overshoot removal, 50 % of testers can correctly point out the quietest one, 50 % does not hear a difference and 0 % answers in the wrong way, confirming the results of the first test and reinforcing the hypothesis that overshoots play an important role in loudness perception.

- When listening to pure tones, testers seem to perceive strong overshoots on valleys (with 550 msec sounds) and, when comparing tones on valleys and peaks, they are most of the time able to state which one is the quietest despite the overshoot behavior. This was tested only through one question since the focus of this thesis was on music, but it is an interesting topic to conduct more research on.

- When evaluating the perceived precision of sounds in each room and with four different sounds, the rank generated by averaging all testers' scores is, most of the times, the same ranking obtained with the Room Slowness and Decay Time parameters. In only two cases two rooms were swtiched (giving the same ranking obtained with Room Inertia), even though in one of the cases the two rooms were only 0.3 % apart.

- Further research should be made to clarify which, if any, among Room Slowness, Inertia and Decay Time is the most significative on the perceived loss of precision and definition after the convolution with a room impulse response. However, it seems that all three parameters are very useful and provide similar results. The reader is reminded that Room Slowness and Inertia provide a better visualization of the temporal behavior of each frequency.

- Running a two-way ANOVA analysis, the type of sound and the room temporal parameters are both relevant for the results, while the interaction between the two variables is not. This means that changing the level of one of those variables has an impact on the result (for example, "slower" rooms tend to generate lower ratings), but that a specific value of one of the variables does not modify the general trend of results (for example, with a particular type of sound, a "slower" room does not generate ratings which are higher than those generated by "fast" rooms).

- The general preference of sound is mostly dependant on the temporal parameters of the room, but also on other factors. The result of question 15,

alongside the specific case scenario of the rooms under comparison, hint that the frequency response curve could be also very important.

- Regarding the words used to describe the test's sounds, it seems that "definito" (italian for well-defined) is the most appropriate word with italian testers to describe dry sounds or sounds convolved with "fast" rooms impulse responses, followed by "preciso" (precise). Sounds convolved with "slow" rooms' impulse responses, which feature higher degradation, are better described by the word "rimbombante" (boomy), followed by the term "risonante" (resonant"). No new significative words were entered.

Chapter 8

Conclusion

8.1 Applications on room acoustics

This thesis work was aimed at studying the psychoacoustic perception of low frequency phenomena in small rooms. In particular, to study the effect and the importance of transient energetic phenomena and to study how the convolution with a room's impulse response impacts the perceived precision and definition of a sound and perceived loudness.

In the following list, all the main results obtained have been summarized. For the specific results of each step of the analysis, the reader is invited to read chapters 5 and 7. The reader is reminded that the term "fast room" is used to describe a room in which the Response Envelopes of the frequency response's peaks are able to reach their steady state even with short bursts, while "slow rooms" is used to describe rooms where the opposite happens.

- Two main behavior and areas of interest emerge from the advanced AQT Room analysis: on valleys of the frequency response, Response Envelopes show overshoot behavior, and this appears to impact the perceived volume of notes whose fundamental is that frequency when varying the note duration. On the peaks of the frequency response, Response Envelopes are quite slow in both reaching the and decaying from the steady state. This appears to act on the perceived definition and precision of notes centered at those frequencies, and on the level perception when varying the note duration (if, for those specific frequencies, Response Envelopes do not reach their actual steady state value for short bursts).

- From the Overshoot Advanced Analysis, it appears that the presence of a valley in the frequency response is correlated to the presence of overshoot behavior.

- The longer the test tones, the more confidently it is possible to state that frequencies reach their actual, theoretical steady state value. For longer bursts, the AQT Steady State Response is the same as the theoretical Frequency Response.

- The Overshoot Response (as defined in the previous chapters) appears to be psychoacoustically more meaningful than the classic Frequency Response when

evaluating short sounds volume perception. In particular, the Overshoot Response developed with 150 milliseconds bursts should be used as it incorporates the effect of rooms with high "Slowness" and "Inertia" values, by featuring a lower value on the frequency response's peaks whose AQT Response Envelope does not reach its theoretical steady state value for really short notes.

- The classic Frequency Response curve appears to be psychoacoustically more meaningful, instead, when evaluating the perceived level of longer sounds.

- Room Slowness and Room Inertia seem to be correlated and feature high values on room modes, while they feature low values on valleys.

- Testers are, for the majority, able to perceive distinctly the precision degradation after the convolution of a sound with the impulse response of a "slow" room, while they are not always able to perceive the precision degradation after the convolution of a sound with the impulse response of a "fast" room. Testers are almost always able to state which room degrades the most the perceived precision of a sound, if directly comparing two rooms with different temporal characteristics.

- While the exact mechanism of pitch perception is not completely known yet, listeners definitely struggle in discriminating the pitch of very short sounds when it is presented non-convolved, convolved on a peak or on a valley. Whether these results arise from a physiological factor alone, or the presence of resonance modes or valleys is actually able to upset the pitch perception for very short sounds, should be object of further studies.

- Given a room mode, and three notes whose fundamental frequency is centered respectively at the same frequency of the room mode, one octave below and one octave above it, the sound which is perceived as being the most resonant is the one with fundamental on the room mode, while the one which is perceived as being the most precise is the third one, with fundamental frequency one octave above the room mode.

- Sounds with a complex spectrum can mask the behavior of a limited number of low-frequency temporal problems, such as frequency with high Room Slowness and Inertia values. Sounds with simpler spectrums (for example, bass notes which feature an harmonic spectrum) seem to mask these problems in a less pronounced way than sounds with more complex spectrum (for example, kick hits, which therefore are more revealing of these type of problems).

- When comparing two short sounds of which one has its fundamental on a room mode, and the other has its fundamental on a valley of the room's frequency response (therefore, showing overshoot behavior), if the overshoot value is similar to the steady state value of the first sound, most testers perceive both sounds as having the same volume. In the same scenario, If the initial and final parts (containing the overshoots) of the second sounds are removed, more testers are able to state that the sound on the valley is quieter than the one on the room mode. This means that the Overshoot Response curve describes the loudness perception for very short notes.

- When comparing directly a sound with overshoot behavior, with the same sound after removing overshoots, a significative percentage of testers can correctly state that the second one is quieter. This confirms the result described in the last point.

- When a short sound is played in a "fast" room, the Overshoot Response seems to be significative for loudness perception, and room modes affect very slightly the perceived precision of the sound.

- When a short sound is played in a "slow" room, the Overshoot Response seems to be significative for loudness perception, and room modes affect heavily the perceived precision of the sound.

- The perceived effect of room modes on precision and definition loss seems to depend mainly on the temporal characteristics of the room (Inertia, Slowness and Decay Time values).

- When listening to pure tones, testers seem to perceive strong overshoots on valleys (with 550 msec sounds) and, when comparing long tones on valleys and peaks, they were most of the time able to state which one was the quietest despite the overshoot behavior.

- When evaluating the perceived precision of sounds in each room and with four different sounds, the rank generated by averaging all testers' scores was, most of the times, the same ranking obtained with the Room Slowness and Decay Time parameters. In only two cases two rooms were switched (giving the same ranking obtained with Room Inertia), even though in one of the cases the two rooms were only 0.3 % apart.

- Running a two-way ANOVA analysis, the type of sound and the room temporal parameters are both relevant for the results, while the interaction between the two variables is not. This means that changing the level of one of those variables has an impact on the result (for example, "slower" rooms tend to generate lower ratings), but that a specific value of one of the variables does not modify the general trend of results (for example, with a particular type of sound, a "slower" room does not generate ratings which are higher than those generated by "fast" rooms).

- The general subjective preference of listening condition seems to be mostly dependant on the temporal parameters of the room, but also on other factors. The frequency response curve can be also very important, especially for musical excerpts or sounds that are long enough to reach their steady state value.

- Regarding Room Slowness, Inertia and Decay Time, it seems that all three parameters are very useful for describing the loss of definition that a sound experience after being convolved with a room's impulse response, and they provide similar results. However, Room Slowness and Inertia provide a better visualization of the temporal behavior of each frequency.

- Regarding the words used to describe the test's sounds, "definito" (italian for well-defined) is the most appropriate word with italian testers to describe

dry sounds or sounds convolved with "fast" rooms impulse responses, followed by "preciso" (precise). Sounds convolved with "slow" rooms' impulse responses, which feature higher degradation, are better described by the word "rimbombante" (boomy), followed by the term "risonante" (resonant"). No new significative words were entered.

8.2 Further possibilities of investigation

Interesting results arise from this analysis, but each of them opens the possibility of investingating in many different directions, in order to confirm these results or introduce even more significative psychoacoustic metrics and concepts. Here are some possible investigation paths:

- These concepts should be tested also in the real world, with listening tests conducted in problematic real small rooms, without having to use headphones and convolution.
- Alternatively, further research should be conducted with synthetic room models in order to separately control some parameters such as Slowness, Inertia and Decay Time, in order to test different scenarios.
- More tests, with non expert listeners, could clarify if these effects are perceived also by non musicians.
- More tests should be conducted in order to confirm that the Overshoot Response is more significative than the Frequency Response when evaluating the level of low-frequency short sounds.
- More research should be made in order to connect the principles of Room Slowness and Inertia to higher order damped systems and the room's surfaces acoustic impedance.
- More tests could be aimed at finding thresholds of detection for most phenomena described in the results, including timbre changes, precision degradation, pitch changes, etc.
- More tests could be aimed at evaluating the effect of stereo sources rather than mono sources.
- Further study should be aimed at verifying which one, if any, between Room Slowness, Room Inertia and Decay Time is the most significative to describe the loss of precision and definition of a sound after its convolution with an impulse response.

Bibliography

[1] URL: http://music.columbia.edu/cmc/MusicAndComputers/chapter3/03_05.php.

[2] URL: http://en.goldenears.net/GR_Headphones/8997.

[3] URL: http://www.phy.mtu.edu/~suits/notefreqs.html.

[4] F. Jacobsen et al. *Fundamentals of Acoustics and Noise Control*. Technical University of Denmark, Department of Electrical Engineering, 2010.

[5] F. Jacobsen et al. *The diffuse sound field: statistical considerations concerning the reverberant field in the steady state*. Lyngby, Denmark: Technical University of Denmark, Acoustic Laboratory, 1979.

[6] P. Schneider et. al. "Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference". In: *Nature Neuroscience* ().

[7] S.E. Olive et. al. "The detection thresholds of resonances at low frequencies". In: *Audio Engineering Society* 45 (1997), pp. 116–128.

[8] Alberto Bellini Angelo Farina Gianfranco Cibelli. "AQT - A New Objective Measurement Of The Acoustical Quality Of Sound Reproduction In Small Compartments]". In: *Audio Engineering Society* (2001).

[9] A. Goldberg B.M. Fazenda M. Stephenson. "Perceptual Thresholds for the effects of room modes as a function of modal decay". In: *J. Acoustic Soc. Am. 137* (2015), pp. 1088–1098.

[10] R. Buchlein. "The audibility of frequency response irregularities". In: (1962).

[11] Ian H. Chan. "Swept Sine Chirps for Measuring Impulse Response". In: *Stanford Research Systems, Inc. 1290-D Reamwood Avenue, Sunnyvale, CA 94089* (2010).

[12] Jamie Angus David Howard. *Acoustics and Psychoacoustics*. Focal Press, 2009.

[13] B.M. Fazenda B.R. Avis W.J. Davies. "Perception of Modal Distribution Metrics in Critical Listening Spaces - Dependence on Room Aspect Ratios". In: *Audio Engineering Society* 53.12 (2005), pp. 1128–1141.

[14] B.M. Fazenda B.R. Avis W.J. Davies. "Thresholds of Detection for Changes to the Q Factor of Low Frequency Modes in Listening Environments]". In: *Audio Engineering Society* 55.7/8 (2003), pp. 611–622.

[15] W.R Fraser D.N. Elliot. *Fatigue and Adaptation, in Foundations of Modern Audition Theory*. New York: J.V. Tobias, Academic Press, 1970.

[16] Robert Erickson. *Sound Structure in Music*. University of California Press, 1975.

[17] Angelo Farina. "Advancements in impulse response measurements by sine sweeps". In: *122nd AES Convention, Vienna* (May 2007).

[18] Angelo Farina. "Impulse Response Measurements". In: *23rd Nordic Sound Symposium, Bolkesjo, Norway* (September 2007).

[19] Angelo Farina. "Simultaneous measurement of impulse response and distortion with a swept-sine technique (preprint 5093)". In: *108th AES Convention, Paris, France* (February 19-22 2000).

[20] Angelo Farina. "Simultaneous measurement of impulse response and distortion with a swept-sine technique". In: *Dipartimento di Ingegneria Industriale, Università di Parma* ().

[21] S.E. Olive F.E. Toole. "The modification of timbre by resonances: perception and measurement". In: *Audio Engineering Society* 36 (1962), pp. 122–142.

[22] Marco Fringuellino. *La Frequenza di Schroeder, SuonoeStudio*. 2007. URL: `http://www.broadcast.it/BeP/Rivista/2007/2007_1/suonoestudio.htm`.

[23] Nastasi Francesco Gabriele Ghelfi Lorenzo Rizzi. "Mixing time measurements in Sabinian and non-Sabinian rooms". In: *Thesis Work - Polytechnic University Of Milan* (2011).

[24] M. Luise G.M. Vitetta. *Teoria dei Segnali*. McGraw-Hill, 2009.

[25] Tor Halmrast. "Musician's Perceived Timbre and Strength in (too) Small Rooms". In: *www.akutek.info* ().

[26] Adam J. Hill MichaelO.J. Hawksford. "Low Frequency Temporal Accuracy of Small Room Sound Reproduction]". In: *AES 133rd Convention, San Francisco, CA, USA* (2012 October 26-29).

[27] Q. Meng D. Sen S. Wang L. Hayes. "Impulse response measurement with sine sweeps and amplitude modulation schemes". In: *School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, Australia* ().

[28] R. Heyser. "Acoustical measurements by time delay spectrometry". In: *Journal of the Audio Engineering Society* 15.4 (October 1967), pp. 370–382.

[29] J.D. Hood. "Studies in auditory fatigue and adaptation". In: *Acta Otolaryngol* 92 (1950), pp. 1–57.

[30] F.Liberatore I.Adami. "La messa a punto del sistema Diffusori-Ambiente". In: *Acustica Applicata srl, via roma 79 Gallicano, Lucca, Italy* ().

[31] P. Laws J. Blauert. "Group Delay Distortions in Electroacoustical Systems]". In: *Journal of the Acoustical Society of America 63 (5)* (May 1978).

[32] Roger George Jackson. *Novel sensors and sensing*. CRC Press, 2004.

[33] Einrich Kuttruff. *Room Acoustics - fifth edition*. Focal Press, 2008.

[34] Leo L.Beranek. *Acoustics*. Amer Inst of Physics, 1986.

[35] Francesco Nastasi Lorenzo Rizzi. "Small studios with gypsum board sound insulation: a review of their room acoustics, details at the low frequencies]". In: *AES 124th Convention, Amsterdam, The Netherlands* (2012 October 26-29).

[36] W. Davies M. Wankling B.M. Fazenda. "The Assessmen of Low Frequency Room Acoustic Parameters using descriptive analysis". In: *Audio Engineering Society* 60.5 (2012, May), pp. 325–337.

[37] Gabriele Ghelfi Michele Ferroni Lorenzo Rizzi. "Evaluation and psychoacoustic validation of techniques for the analysis of low frequency resonance modes in real small rooms". In: *Thesis Work - Polytechnic University Of Milan* (2015).

[38] Brian C.J. Moore. *An introduction to the psychology of hearing*. Academic Press, fourth edition, 2001.

[39] Philipp Newell. *Recording Studio Design*. Focal Press, 2013.

[40] A.M. Noxon. "The Music Articulation Test Tone (MATT)". In: *Acoustic Sciences Corporation (ASC), 4275 West Fifth Avenue, Eugene, Oregon 97440 - U. S. A.* ().

[41] R. Plomp. *Timbre as multidimensional attribute of complex tones, in Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, 1970.

[42] Tomas Salava. "Imperfections at Low Frequencies - how much are they audible or annoying?" In: *AES 116th Convention, Berlin, Germany* (2004 May 8-11).

[43] B. Scharf. *Loudness adaptation, in Hearing Research and Theory*. New York: Academic Press, 1983.

[44] J.F. Schouten. "The Perception of Timbre". In: *Reports of the 6th international Congress on Acoustics, Tokyo* (1968).

[45] M. R. Schroeder. "New method of measuring reverberation time". In: *Journal of Audio Engineering Society* (1965).

[46] M.R. Schroeder. "Schroeder Frequency Revisited". In: *J. Acoust. Soc. Am.* 99 (1996), pp. 3240–3241.

[47] Soren Bech Slawomir Zielinski Francis Rumsey. "On Some Biases Encountered in Modern Audio Quality Listening Tests - A Review". In: *Audio Engineering Society* 56.6 (2008), pp. 427–451.

[48] Archambeau Dominique Stan Guy-Bart Embrechts Jean-Jacques. "Comparison of different impulse response measurement techniques". In: *Institut Montefiore B28, Sart Tilman, B-4000 LIEGE 1 BELGIUM* (2002 December).

[49] Albert Bregman Stephen McAdams. "Hearing Musical Streams". In: *Computer Music Journal* 3.4 (1979), pp. 26–43.

[50] Floyd E. Toole. *Sound Reproduction - Loudspeakers and Rooms*. Spon Press, 2009.

[51] Matti Karjalainen Poju Antsalo Aki Makivirta Vesa Valimaki. "Perception of Temporal Decay of Low Frequency Room Modes]". In: *AES 116th Convention, Berlin, Germany* (2004 May 8-11).

[52] Simeon Delikaris-Manias Vincent Koehl Mathieu Paquier. "Comparison of subjective assessments obtained from listening tests through headphones and loudspeakers setups]". In: *AES 131st Convention, New York, USA* (2011).