A HYBRID METHOD FOR IMPULSE RESPONSE MEASUREMENTS WITH SYNTHESIZED MUSICAL TONES AND MASKED-MLS STIMULI

Massimo SERAFINI

Ph.D. Thesis **2015**

A HYBRID METHOD FOR IMPULSE RESPONSE MEASUREMENTS WITH SYNTHESIZED MUSICAL TONES AND MASKED-MLS STIMULI

Massimo SERAFINI

Acoustics Research Centre

School of Computing, Science and Engineering

University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy, December 2015

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deep gratitude to my supervisor, Dr. Francis Li, for his inspiration and help in my research, and encouragement and reassurance during examination and conference presentations. His advice has been fundamental for overcoming all the difficulties I have encountered during this period of study. I want also to express my appreciation to my co-supervisor, Prof. Trevor Cox, for his support and supervision during the writing up of the thesis.

I would also like to thank all my colleagues and the academic staff within the Acoustics Research Centre at the University of Salford, with whom I have had the pleasure of sharing an office, time, food, and a billion cups of tea. Numerous prolific scientific discussions that I have had with many of them throughout my stay at Salford have enriched my researching skills, and motivated me to pursue and complete this PhD.

In particular, I want to thank my desk-neighbours, James Woodcock, Eulalia Peris, Rodolfo Venegas, Jenna Condie, and my country mate, Gennaro Sica. Special thanks goes to my friend and English teacher, Helena Stakounis, whose teaching skills have helped me to learn and love the English language.

In addition, I am particularly thankful to Francesca Lupo, who helped and motivated me to start this academic journey in Salford and to Aminah Barnes, to whom I will forever be grateful for the shared classes, trips, and countless meals.

Finally, I express my appreciation for my parents, brother and sister, and all the friends in both the UK and Italy, who have constantly supported, cheered me up, and loved me throughout this time and to my grandparents who are always by my side.

I would also like to acknowledge the support of the Leverhulme Trust, for funding this research project and my scholarship, and the RAI (RadioTelevisione Italiana) for giving me such a long sabbatical time off work to attend and complete this PhD.

I want also to express my sympathy to Philip Kamara, a humble man, a real friend who never be forgotten. You fought many hard battles during all your life, and you are now facing the hardest of them all. My most sincere and deep hopes that you will win this one too.

List of Abbreviations

IR	Impulse Response
PIR	Periodic Impulse Response
RIR	Room Impulse Response
TF	Transfer Function
FRF	Frequency Response Function
RFR	Room Response Function
LTI	Linear Time Invariant
FT	Fourier Transform
IFT	Inverse Fourier Transform
RT	Reverberation Time
EDT	Early Decay Time
ITDG	Initial Time Delay Gap
D_{50}	Definition Index
C_{80}	Clarity Index
STI	Speech Transmission Index
LF	Later Fraction
IACC	Inter-Aural Cross-Correlation
EDC	Energy Decay Curve
ETC	Energy-Time Curve
JND	Just Noticeable Difference
JNND	Just Non Noticeable Difference
DL	Difference Limens
SNR	Signal-to-Noise Ratio
CF	Crest Factor
RMS	Root Mean Square
MSE	Mean Square Energy
ACF	Auto Correlation Function
CCF	Cross Correlation Function
PSD	Power Spectral Density
CPSD	Cross Power Spectral Density
TDS	Time Delay Spectrometry

FFT	Fast Fourier Transform
FHT	Fast Hadamard Transform
PRBS	Pseudorandom Binary Sequence
MTS	MIDI Tuning Standard
MLS	Maximum Length Sequence
MLSSA	Maximum Length Sequence System Analyser
IRS	Inverse Repeated Sequence
TSP	Time-Stretch Pulse
ESS	Exponential Sine Sweep
LSS	Linear Sine Sweep
MLE	Maximum Likelihood Estimation
ASR	Automatic Speech Recognition
FIR	Finite Impulse Response Filter
DSP	Digital Signal Processing
PA	Public Address System
DC	Direct Current
ATH	Absolute Threshold of Hearing
THQ	Threshold in Quiet
ELLC	Equal Loudness Level Contours
CF	Characteristic Frequency
CB	Critical Band
ERB	Equivalent Rectangular Bandwidth
SMR	Signal-to-Mask Ratio
NMT	Noise Masking Tone
TMT	Tone Masking Tone
TMN	Tone Masking Noise
NMN	Noise Masking Noise
MPEG	Motion Picture Experts Group
MP1	MPEG Layer 1
NR	Noise Rating
PNC	Preferred Noise Criterion
L_{Aeq}	Equivalent Continuous Noise level (A-Weighted)

Abstract

Impulse responses or transfer functions are descriptions of acoustic and audio transmission channels, which completely characterise point-to-point propagation of sound in room acoustics or input-output relationships of linear systems in electroacoustics. Measurements of impulse responses or transfer functions are routinely carried out first to determine critical room acoustics parameters of enclosures such as concert halls, theatres and auditoria, or technical specification of electroacoustic transducers, i.e. loudspeakers and microphones. In room acoustics measurements, tradition techniques employ noisy testing signals as probe stimuli, which are unpleasant and intolerable to audiences. This hinders occupied measurements to be taken in many cases. Predicted in-use parameters from unoccupied measurements are known to be unreliable and problematic. It is also well appreciated in room acoustics research community that the use of musical or music-masked probe stimuli can mitigate problems of occupied measurements. It is therefore hypothesised as a starting point of this thesis that the use of musical tone like stimuli or musically masked testing signals can be used to determine impulse responses or transfer functions.

Based on the above hypothesis, this thesis develops a new hybrid technique, in which narrow band linear chirps, called “presto-chirps” centred on musical notes are used to measure impulse responses in low to mid frequency bands, and music-masked maximum-length-sequences are deployed to obtain those in higher frequency bands. Broadband impulse responses are then obtained by combining the measured lower and higher frequency impulse responses.

To test the hypothesis and identify the potential and limitations of the developed technique, mathematical formulation and analysis, computer simulations and real room measurements have been carried out and documented in this thesis.

Investigation results show that purposely tailored and windowed narrow chirps that emulate musical tones can be used as probe stimuli to measure impulse responses or transfer functions with an uncompromised accuracy. It is found that Hanning windows are almost optimal for this application. This method covers frequency ranges commonly quoted in room acoustics investigations. Music-masked maximum-length sequences are found to be able to obtain in impulse responses or transfer functions in higher frequency. However, if completely masked stimuli are sought, the resulted signal to noise ratios in the measurements is limited, or the required averaging is going to be overly prolonged. Nevertheless, the masking music can still potentially be used as a distracter to make the audience more forgiving to the hissing noise from maximum length sequences, facilitating the occupied measurements.

1 Introduction

It is well known that the presence of an audience may alter the acoustics of enclosed spaces considerably. The current knowledgebase about suitable acoustics in diverse spaces and design guidelines for such spaces are acquired and derived predominantly from unoccupied data. The estimations made from unoccupied data are known to be problematic in some cases. Occupied measurements are therefore sought after to better understand acoustics under in-use conditions.

1.1 Motivations

High definition audio reproduction, e.g. surround sound and spatial audio techniques (Dolby, DTS, Ambisonics, etc.) rely heavily on the listening environment and positions. For in situ measurements, it may be useful being able to determine impulse responses from loudspeakers to listeners so that each channel can be individually calibrated and room effect compensated for. A smart phone powered with appropriate Apps and Bluetooth communication links can provide simple and economic in situ acoustic measurements for this purpose.

Moreover, performing a great number of occupied measurements might improve computer simulations of architectural acoustics, since it could provide accurate data about sound absorption caused by an audience distributed in diverse geometric configurations.

In conclusion, acousticians, sound engineers and ultimately, the audience and general public might benefit from a method able to perform impulse response measurements in occupied spaces.

However, the biggest challenge in developing methods for occupied measurements is to identify suitable probe signals. Common impulse response measurement are carried out using broadband deterministic signals, such as the exponential sine sweep (ESS) as probe stimuli at a high sound pressure level, which can be unbearable to occupants. Maximum length sequence (MLS) is also commonly used for audio measurements, and despite its intrinsic noise rejection property that can theoretically allow inaudible measurements to be carried out, overly prolonged averaging makes the measurement susceptible to time variance and hence unreliable. Conversely, naturalistic sounds like speech and music, which can be use within a dual channel FFT technique (2chFFT), although being straightforward stimuli for occupied measurements, often lack frequency fullness and sufficient signal-to-noise ratio (SNR) in certain sub-bands.

Similarly, blind and semi-blind techniques, which were found able to estimate some acoustic parameters of simply-shaped spaces using speech (Li & Cox, 2003; Li, 2005), and music (Li et al., 2007; Kendrick et al. 2007; 2008), did not solve the problem of broadband impulse response measurements.

1.2 Proposed Solution

To overcome the limitations of the aforementioned methods in dealing with occupied measurements, this thesis proposes the combined use of two novel impulse response measurements techniques to tackle the problem. Each method has been designed to measure impulse responses in a specific and complementary frequency range. The two methods aim at mitigating the disturbance to occupants and thus enable occupied in-situ measurements.

- The first method uses semitone-width chirps constructed based on the frequencies of notes in the equal tempered scale, and used to generate musical stimuli. The method is structured to measure the impulse responses from 27.5 Hz up to 4.69 kHz to mirror the frequency range (fundamental frequencies) of a grand piano (plus one semitone).

- The second method, which has been developed to measure impulse response from 4.69 kHz to 20 kHz, uses maximum length sequences, masked by music, used to perform unnoticeable measurements. The masker music is low pass filtered at 4.69 kHz, to increase the signal to noise ratio in higher bands, and then allowing MLS measurements to be completed in a short time, therefore limiting time variance susceptibility.

The final step is to incorporate the two novel techniques into a single measurement framework, which gives birth to the “Hybrid method” presented and discussed in this thesis.

1.3 Aim and Objectives

The overall aim of this study is to explore if deterministic but musical probe stimuli can be used to accurate determine room impulse responses or transfer functions, and to develop associated signal-processing algorithms.

The objectives of this project are to

(1) Develop a linear system identification method using a series of narrow-band chirp signals (presto chirps) centred around equal temperament chromatic scales,
(2) Develop algorithms to mitigate artefacts caused by fading in and out so that accurate impulse response can be obtained,
(3) Critically evaluate performance of the presto-chirp method to validate its applicability in the room acoustics measurement context and identify its limitations,
(4) Derive a masking model and develop a method to measure the impulse responses in higher frequency sub-bands with maximum length sequences masked by music,
(5) Critically evaluate and validate the masked MLS method, and
(6) Combine the presto-chirps and masked MLS methods to enable measurement of broadband impulse responses.

1.4 Thesis Structure

The thesis is structured as follows:

- *Chapter 1* contains an introduction of the PhD research project. It also presents the motivations, aim and objectives of the research. It finishes by presenting an outline of the thesis.

- *Chapter 2* discusses existing measurement methods and techniques used to determine room impulse responses. It also includes a review of blind and semi-blind estimation methods and culminates with an outline of the proposed new solutions to occupied measurements.

- *Chapter 3* discusses the presto-chirps measurement method, starting with its mathematical formulation and theoretical analysis. It follows by detailing the signal processing methods with an emphasis on methods to mitigate errors and artefacts. The chapter ends by presenting the implementation of a musical stimulus.

- *Chapter 4* validates the presto-chirps method by computer simulation and in situ measurements.

- *Chapter 5* discusses the masked-MLS measurements method. It introduces some basic psychoacoustic theories and outlines how a “lower” masking thresholds is estimated from masker music, using the psychoacoustic model

implemented in the MPEG1 layer 1 codec. It concludes by showing an example of estimations of measurement time based on the MLS level and initial signal-to-noise ratio.

- *Chapter 6* shows how the measurement scheme used to perform masked-MLS measurements has been structured. It also presents the investigations results of an in situ measurement in a studio room.

- *Chapter 7* explains the hybrid method and shows how the impulse responses measured using the presto-chirps and masked-MLS methods are combined.

- *Chapter 8* presents a summary, draws up conclusions, and proposes further work.

- Appendix A presents an example of how presto-chirps can be adapted, based on an existing song, to perform accurate acoustic measurements.

- Appendix B discusses further uses of the presto-chirps.

- Appendix C presents two further cases study about in situ measurements carried out using the presto-chirps method.

2 Literature Review

The modern era of architectural acoustics is deemed to have been started in the late 1800's/early 1900's by Sabine (1923). He first studied sound decays in enclosed spaces as functions of the frequency and amount/distribution of acoustic absorbent materials, using only a stopwatch, pipe organs, and his ears as measuring instruments. He also introduced the concept of reverberation time (RT) and the first equation that relates the RT to the volume of a space and the total acoustic absorption. This new scientific discipline inspired other researchers to investigate further the physics of sound propagation and the relationships between objective and measurable physic quantities and their associated objective evaluation of acoustic quality. In doing so, the RT has always been used as a main descriptor, and Sabine's original equation had been rewritten and "corrected" in the following year by other researchers (Eyring, 1930; Millington, 1932; Sette, 1933; Fitzroy, 1959). All these pieces of work are theoretical derivations of the RT based on the geometry of the spaces and the distribution, and properties of acoustic materials and their specific acoustic absorption coefficients.

Measurements of RTs became fundamental to characterise the acoustic quality of existing spaces. Sabine's approach to measure RT from the sudden stop of the pipes' sounds requested to perform single measurements at different frequencies (he only used five notes). Such measurements were not particularly accurate compared with today's standards, but considering the lack of technology at that time, they were both intuitive and straightforward.

With the advent of (audio) technology and the developing of loudspeakers, microphones, personal computers, along with the theories and principles of sound processing (an exhaustive description can be found in Oppenheim & Schafer (2009)), acoustic measurements changed their pace to a modern acoustic analysis. Broadband stochastic noise replaced the organ's notes as probe stimuli, and although reverberation time was still measured based on the sound decay after the switch-off of the sound source more accurate electronic devices replaced the stopwatch. This measurement method is known as the "interrupted noise method" and is described in the standard BS EN ISO 3382-2:2008. A limitation of this kind of measurement is that, due to the randomness of noise, many averages have to be taken to achieve a smooth and accurate sound decay, from which the RT is calculated.

2.1 Backward Integration and Room Impulse Responses

In the late sixties, however, Schroeder (1965) developed a new methodology for measuring reverberation times without using an ensemble average. He demonstrated that backward integrating the response of a system to an impulsive excitation is mathematically equal to performing the ensemble average of several noise-interrupted decays. The resulting Schroeder integral curve, commonly called energy decay curve (EDC) and also referred to as energy time curves (ETC) in Li & Lam, (2005), and is mathematically defined as

$$EDC(t) \stackrel{\text{def}}{=} \int_t^\infty h^2(t)dt = \int_0^\infty h^2(t)dt - \int_0^t h^2(t)dt \qquad (1)$$

where $h^2(t)$ is the measured impulse (IR) response, squared, and *t* is the time variable. In Figure 1 is shown an example of the application of the Schroeder's formula. The RT is then estimated upon a trend line of the Schroeder integral curve (in dB scale).

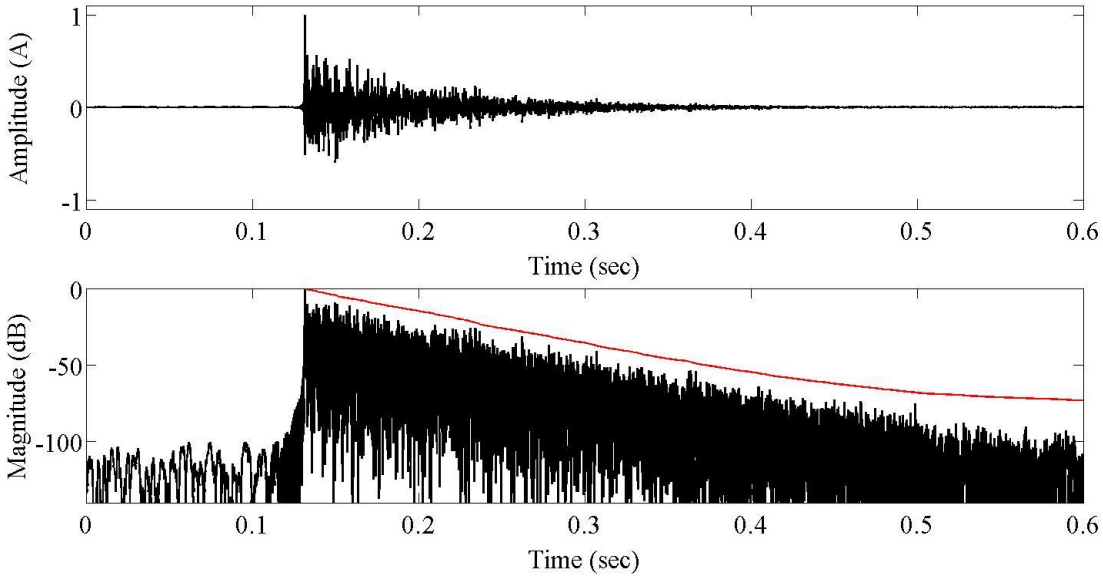


Figure 1 – Example of a Schroder backward integration (red line).

The response of a room to an impulsive excitation is called the room impulse response (RIR), which practically is the time signature of an "impulsive sound" emitted by a sound source and travelling in a room to a receiver position. The geometrical shape of the listening environment affects and defines the distribution of the sound in the enclosure and therefore the sound that reach the listener in a specific position. It is common to define the sound arriving at the listener position into three distinct stages; direct sound, early reflections, and reverberation. The direct sound is related to the sound travelling directly from the loudspeaker to the listener without bouncing on other surfaces. Conversely, the early reflection stage is characterised by the sound reaching the listener after being reflected by surfaces nearby the source and the receivers. The reverberation stage is characterised by reflections coming from all

directions, which are therefore impossible to be tracked back to the source. Figure 2 shows these three stages, along with a schematic of sound travelling in a room, from a loudspeaker to a listener.

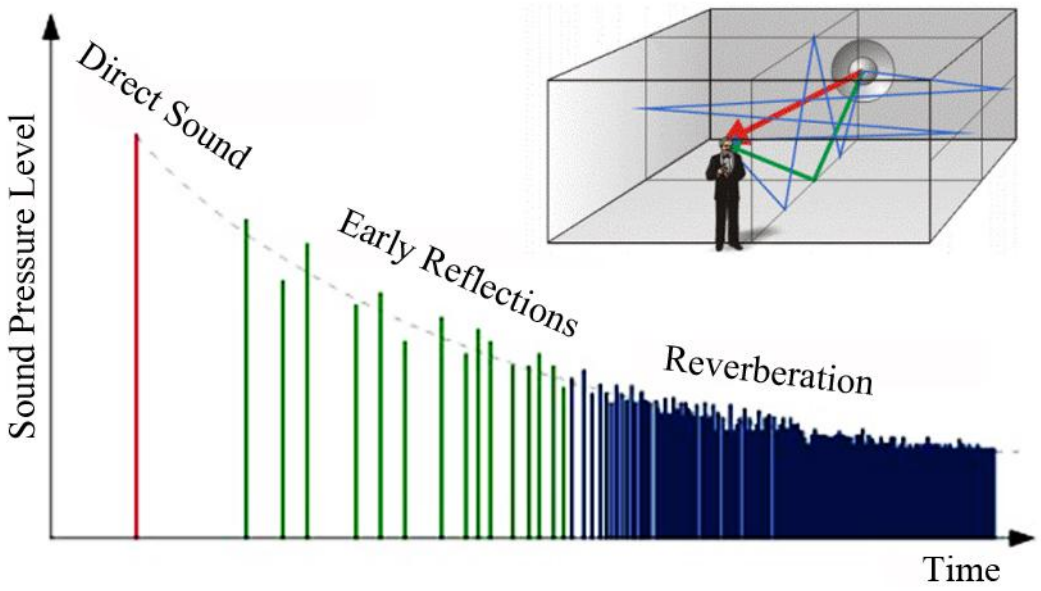


Figure 2 – Schematic of the propagation of sound within an enclosure.

2.2 Acoustic Parameters

Acoustical spaces are qualitatively characterised by measuring several acoustic parameters, which are calculated from their RIRs. Acoustic parameters are used to correlate objective physical quantities with subjective impressions of acoustic quality. Indeed, the knowledge of those quantities allows for determining which specific use an acoustical space will best suit, i.e., music, opera, speech, general purpose, and so on. Similarly, when new places are constructed (houses, schools, hospitals, theatres, concert halls, and so on), their acoustical characteristics should reflect their purposes and uses. However, As pointed out by Hak et al, (2012), the correct estimation of acoustic parameters is only achieved when measures lead to a sufficient and clear

energy decay curve. Background noise, changes of environmental characteristics (humidity, temperature, etc.), and non-linear harmonic distortions introduced by the measurement chain, in addition to the energy of the testing signal, all contribute to the actual signal-to-noise ratio. A high SNR results in a wider noise-free EDC, where RT and other acoustic parameters can be calculated. Table 1 reports the mathematical formulae of the acoustic parameters used within this thesis to estimate the accuracy of the presented measurement methods. A more complete list and description of acoustic parameters can be found in Bradley (2011), and Hak et al (2012), in addition to the BS EN ISO 3382-1:2009 standard.

- Clarity (C_{80}) [dB]

$$C_{80} = 10\log_{10}\frac{\int_0^{80ms} h^2(t)dt}{\int_{80ms}^{\infty} h^2(t)dt} \ [dB] \quad (2)$$

The Clarity Index, introduced by Reichardt et al. (1974) is calculated as the ratio of the energy of the impulse response ($h(t)$) in its first 80ms and its energy from 80ms to infinity, measured in dB.

- Definition (D_{50}) [%]

$$D_{50} = 100\frac{\int_0^{50ms} h^2(t)dt}{\int_{0ms}^{\infty} h^2(t)dt}\% \quad (3)$$

Similar to the clarity index but it is calculated as the percentage ratio of the energy of the impulse response ($h(t)$) in its first 50ms and its energy from 50ms to infinity.

- Reverberation Time RT_{60} $[s]$

$$L(t) = 10\log_{10}\frac{\int_t^{\infty} h^2(t)dt}{\int_0^{\infty} h^2(t)dt} \ [dB] \tag{4}$$

The RT_{60}, proposed by Sabine, was the first acoustical parameters ever used. It is related to the time needed for the sound pressure level of a stationary sound to decrease of 60 dB, once it is turned off. This parameter is not widely used due to the difficulty of achieve such a wide noise free dynamic range in acoustic measurements. However, due to possibility of achieve a great SNR using chirps signals, it is used in this thesis to mark the robustness of the measurements.

- Reverberation Time RT_{30} $[s]$

$$L(t) = 10\log_{10}\frac{\int_t^{\infty} h^2(t)dt}{\int_0^{\infty} h^2(t)dt} \ [dB] \tag{5}$$

Similarly to the RT_{60}, the RT_{30} is the measure of the time the energy of the impulse response drop, whit the difference that it is calculated in the range from -5 to – 35 dB of the least-squares fit of the decay curve. Starting from -5 dB mitigates estimation errors due to the direct sound and strong early reflection.

• Early Decay Time EDT [s] $$L(t) = 10 \log_{10} \frac{\int_t^\infty h^2(t)dt}{\int_0^\infty h^2(t)dt} \ [dB] \quad (6)$$ The early decay time is the measure of the time needed for the sound to the decay of 10 dB (and then multiplied by 6 to relate it to a -60 dB decay). It is calculated along the Schroeder's plot, which is turn is calculated from the energy decay. The EDT, as discussed in Gade (1994), closely related to the "reverberance" of a room and it is nowadays commonly used to characterise room reverberation, especially when used for music.

Table 1 – Mathematical, and descriptive, definitions of the acoustic parameters used in this thesis.

2.3 The Theory of Linear and Time-Invariant Systems

When measuring the room impulse responses, the room, the source, and the receiver are treated together as a linear system. The theory of linear and time-invariant (LTI) systems is a part of applied mathematics concerning the definition and properties of a specific class of systems, i.e. linear and time-invariant. These systems have a special connection to acoustics because they can represent acoustical spaces under some specific assumptions. A linear system is a system that abides by the mathematical additive and homogeneity properties, thus to the superposition principle. The additive property states that the linear response of a system to individual inputs is equal to the system's response to the sum of those inputs

$$F\{x_1(t)\} + F\{x_2(t)\} + \cdots + F\{x_n(t)\} = F\{x_1(t) + x_2(t) + \cdots + x_n(t)\} \quad (7)$$

where $F\{\}$ is the linear function operator, and $x_n(t)$ are the inputs. The homogeneity property instead implies that a multiplication/division of an input signal by a scalar (an increase/decrease of the input level, for example) produces the same alteration to the output of the system

$$F\{\alpha x(t)\} = \alpha F\{x(t)\} \quad (8)$$

where α is the scalar. Another fundamental property of LTI systems is the time-invariance, which specifics that the response of a system is independent to time. Thus, if the input of a system is time-shifted, the output of that system will be time-shifted by the same amount.

Time invariance and linearity are ideal attributes of acoustic spaces, but in real case scenarios, although rooms are linear systems they are also generally time variant. Since time invariance is only guaranteed when the physical state of a system under test does not vary during the measurements, prolonged averaging should be avoided. Conversely, linearity is unlikely to be a feature of the electroacoustic devices used in the measurements themselves; therefore, methods able to reject non-linearity are preferred.

2.3.1 Impulse Responses

The fundamental property of LTI systems is that they are completely defined by their impulse responses. Thus, by inputting an "ideal" impulse (Dirac delta function) into a system, the output, the response of the system to that impulse, will be a unique and characteristic function that completely describes the system. Mathematically, this is described by $F\{\delta(t)\} = h(t)$, where $\delta(t)$ is the Dirac delta function and $h(t)$ is the impulse response of the system. Mathematically the Dirac delta function is defined as

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ 1, & t = 0 \end{cases} \tag{9}$$

and in its digital form, where it is commonly named as unit impulse, is

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \tag{10}$$

where $\delta[n] = \delta(nT)$ is the discrete time version of $\delta(t)$ sampled every $T = 1/fs$. Knowing an IR of a system allows for determining its output in response to any input, thus the output is the convolution of the input signal and the IR of the system, which is mathematically defined as

$$y(t) = x(t) * h(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)dt \tag{11}$$

where the symbol $*$ represents the convolution operator, and $y(t)$ is the output of a system with impulse response $h(t)$ due to an input signal $x(t)$. Eq. (11) is commonly called the convolution integral. The discrete version of the convolution used for digital applications is instead

$$y[n] = x[n] * h[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{+\infty} x[m]h[m-n] \tag{12}$$

2.3.2 Transfer Functions

Besides the time signature of a system, its frequency distribution is also a distinctive trait worthy of analysis. The time and frequency domains represent two different points of view (temporal and spectral) of the same system, and the knowledge of one means the knowledge of the other. In fact, through the Fourier transform pairs, it is possible to "swap" from one domain to the other. Mathematically, a Fourier transformation (FT) applied to the IR of system gives the so-called transfer function (TF) of that system. Conversely, the inverse Fourier transformation (IFT) can be used

to "reverse" the domain shift. These Fourier transformations are mathematically represented by the following equations,

$$\begin{cases} \mathcal{F}\{h(t)\} = H(j\omega) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} h(t)e^{-j\omega t}dt \\ \mathcal{F}^{-1}\{H(j\omega)\} = h(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} H(j\omega)e^{j\omega t}d\omega \end{cases} \tag{13}$$

where $\mathcal{F}\{\,\}$ and $\mathcal{F}^{-1}\{\,\}$ are the FT the IFT respectively. $H(f)$ is referred to as the frequency response function (FRF) of the system. Figure 3 reports a diagraph of the relation between the input, the impulse response, and the output and their associate Fourier transforms.

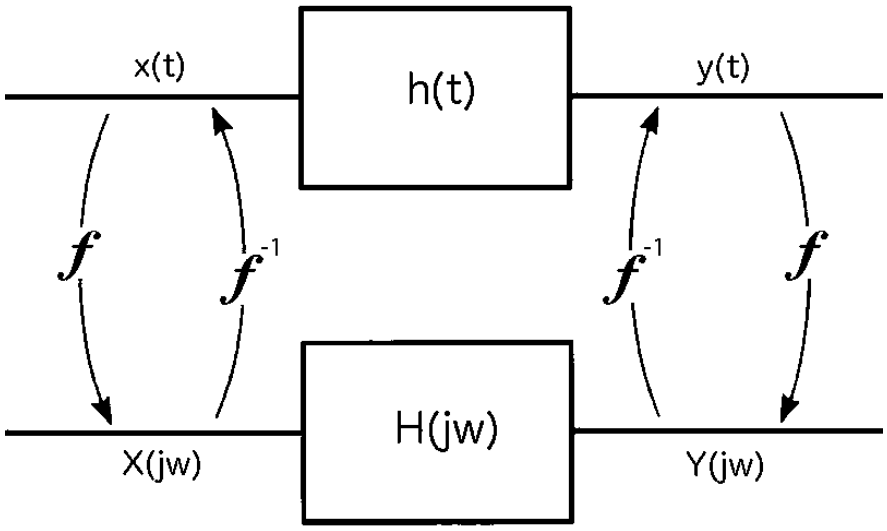


Figure 3 – Schematic of the dualism between time and frequency domain, and their relation with the Fourier transform.

2.4 Measurements of Room Impulse Responses

After the Schroeder's backwards-integration technique, the measurements of impulse responses have become a fundamental step to provide accurate measure of RTs, and the other acoustic parameters. The advent of electroacoustic devices and computers, allowed new RIR measurement methods to be developed that could provide more accurate and reliable measurements. This paragraph reviews the principal measurement methods used to measure both the RIRs and the FRFs of acoustical spaces.

2.4.1 Impulsive Excitations

In the dawn of RIR measurements, the earliest substitutes for ideal impulses were impulsive noises generated from gunshots or popping balloons. However, the loud and unpredictable nature of such signals produces uneven excited frequency spectra and a low and non-uniform SNR, which may result in large errors. Moreover, pistol (for room measurements) and loud sounds are not very pleasant. However, with the introduction of loudspeakers, microphones, and signal generators/personal computers, impulses could be digitally generated, thus removing the downsides of the "manually" generated pulse. However, when a single pulse is used, it has a high crest factor (CF) and this significantly reduces the achievable SNR for a single pulse, averaging is generally used to increase it. The CF is defined as the ratio of the peak amplitude value of the signal (Peak) by the root mean square (RMS) of the signal, thus $\mathrm{CF} = Peak/RMS$. It is a dimensionless quantity and it is generally expressed in dB, especially in acoustic applications. The CF gives a ready measure of the amount of energy that a signal can transfer to a system. For example, the energy of an impulse is

concentrated in a short amount of time and has a high CF. Conversely, the energy of a steady signal, like the direct current (DC) or a square wave, is spread over a longer time interval, and has a CF of 0 dB (the lowest possible).

2.4.2 Discrete Tones

The manufacturing and development of electroacoustic devices implied that tests and measurements were needed to characterise their performances. Pulse or mono frequency signals were commonly used for such measurements. For example, in a method known as the "stepped tone method", mono-frequency signals are input into the systems/rooms and the outputs are analysed through the fast Fourier transform (FFT). The frequency is varied in successive steps by a predefined quantity, which could be linear or logarithmic. This approach was ideal for testing non-linearity and "near" resonance frequencies, since the frequency, the phase and the amplitude of the tone can be easily varied. Using tones generally results in a high signal to noise ratio. Variations of the method include the use of two simultaneously played tones, which can be used for measuring intermodulation distortion and difference frequency distortion. Playing together several tones leads to a multitone signal which is a way to speed up the otherwise time consuming measurement of broadband systems using the single tone approach. The advantage of such stimuli is that they can cover a broader frequency range in just one measurement; however, this comports the impossibility of measuring systems' non-linearity. Information about tone-based measurements can be found in Schoukens et al. (2000).

2.4.3 Time Delay Spectrometry

In the 1967, Heyser introduced a new measurement technique named the time delay spectrometry (TDS), where a linear frequency varying sinusoid, thus a sine sweep (also called chirps), is used as a probe signal to measure the frequency response of loudspeakers in reverberant rooms. Besides Heyser, several other authors have carried out studies on the theory of TDS (Vanderkooy, 1985; Poletti, 1988; Cook & Zaknich, 1996), and the method was employed for measurements of acoustic and electroacoustic systems for a long time.

In practice, a sine sweep is fed into the system under test and the recorded output is multiplied by both the original sine sweep (which gives the system's transfer function real part) and its 90° phase-shifted cosine (which gives the system's transfer function imaginary part). This product gives sum and differences of the input frequencies. A frequency-variable fixed-bandwidth low pass filter, offset in respect to the generated signal by a predetermined amount, is used to filter out the sum terms and let the difference terms pass. The cut off frequency of the filter increases linearly with respect to time with the same ratio of the probe signal. When the stimulus and the sine generator have same instantaneous frequency, their difference sum will be near zero and will not be attenuated by the low pass filter. On the other hand, the travelling sound reflected by the room will take a longer time to arrive at the receiver position, and this will result in a higher frequency difference term, which will then be attenuated by the filter. By carefully setting the filter initial offset and bandwidth, it is possible to achieve quasi-free-field measurements, since all the reflections, the distortions, and noise (the part above the cut off frequency) will be suppressed by the

filter. Another positive feature of TDS is its robustness against time variance and the high achievable signal to ratio that also takes advantage of its optimal 3dB crest factor.

On the other hand, TDS probe has a white spectrum and may feed little energy at low frequency bands, which are usually more strongly affected by the background noise. However, using two linear sweeps to perform the measurements, one for the low frequency region and another for the high frequency region could be a worthwhile solution to achieve higher SNR in the low region of the audio spectrum. Another drawback is inherent to the sudden onset of the sweep, which creates a rippling effect in the initial and final part of the spectrum of the sweep. A way to circumvent the problem could be to let the sweep start and end at frequencies outside the range of interest (Müller & Massarani, 2001) and then filter the result.

2.4.4 Maximum Length Sequence

During the sixties a special class of pseudo-random binary sequences (PRBS), namely maximum length sequences (MLS), were investigated by Davies (1966), who showed that the MLS autocorrelation function (ACF) yields to an almost perfect Dirac delta pulse. The use of MLS for acoustic measurements was first experimented by Schroeder (1979), who showed the advantage of such a measurement method over impulsive noise based measurements. Following studies carried out by Borish & Angell, (1983), and Rife & Vanderkooy, (1989), have further contributed to the MLS reliability and efficiency. The advent of MLS into acoustic practices is mainly due to Rife (1987), who first implemented the method for a personal computer based analysis system called MLSSA (maximum length sequence system analyser). One

important features of the MLS is its great noise rejection to all kinds of noises, which is due to the phase randomisation of non-coherent signals that are spread after the deconvolution along all the measured impulse responses (Nielsen, 1997). This important effect yields to the possibility of using MLS even within negative SNR, which may be used when supported by an averaging technique to experiment on inaudible or unobtrusive acoustic measurements. In fact, Schroeder (1979) reported he was able, using audible low-level MLS, to measure the early part of the RIRs of some occupied classrooms during his lectures. Details of measurements have not been given but he proved that under some circumstances the MLS is an interesting candidate to be used for occupied measurements. Nonetheless, for more precise measurements of bigger halls, where time variance can be consistent and lengthy averages are needed, MLS has been proven ineffective in delivering accurate measure of RIRs (Vorlander & Kob, 1997; Svensson & Nielsen, 1999; Satoh et al., 2002). Moreover, measurements carried out in occupied conditions would also be further complicated by the presence of an audience, which will likely increase the halls' time variability. In fact, as pointed out by some authors (Li, 2002; Li & Serafini, 2010) , if measurements are taken using an MLS 30 dB below the level of the music, for example, the overall measurement duration will be as long as 11 hours (Serafini et al., 2010). That would only be feasible in the unlikely hypothesis that time variance and audience will be unchanged during the whole measurement time.

2.4.4.1 Theoretical Aspects of MLS

MLS are discrete time periodic sequence (Rife & Vanderkooy, 1989), which length, in samples, is equal to $L = 2^n - 1$ (the exponent n also gives the "order" of the sequence). MLS are generated, according to primitive polynomials (available in

literature or in the web), using digital feedback shift registers. Practically, the shift registers cycle through every possible binary values (or "states") of the sequence (from this property came the notation "maximum") with the exception of a vector with all zeros. A schematic of an MLS generator line is shown in Figure 4. The XOR component is a modulo-2 sum operator, which is used to realise the feedback line.

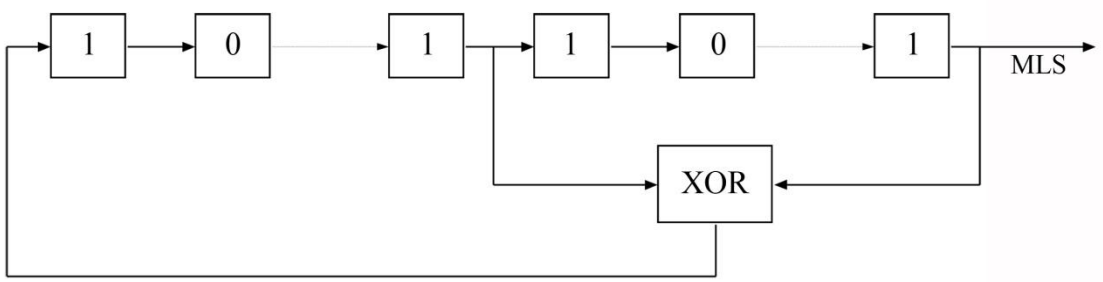


Figure 4 – An example of the generation of an ML sequence using shift registers.

The MLS power spectral density (PSD) is white, but conversely to white noise, the periodicity of MLS results in having a discrete set of frequency points excited with the same amount of energy in just one complete period (Rife & Vanderkooy, 1989). The density and frequency of those points depend both on the length of the sequence and on its sampling frequency, following the relation $f_s/2L$, thus the sampling frequency dived by twice the number of samples. For practical measurements, the MLS is commonly used within its symmetrical representation, which is obtained substituting every zero with one, and every original one with minus one. MLS (periodic) autocorrelation is

$$\Omega_{xx}[n] = \frac{1}{L}\sum_{k=0}^{L-1} x'[k]x'[k+n] = x'[n] \circledast x'[n] \tag{14}$$

where the symbol ⊛ represents the circular/periodic convolution operator, and $x'[n]$ is the periodic (modulo N) MLS sequence, and the factor $\frac{1}{L}$ is a small DC offset. Rife & Vanderkooy (1989) postulated that it is far more convenient to use a normalisation factor of L+1 in place of L. This permits to express the MLS' autocorrelation as a sum of periodic unit-sample pulses ($\delta'[n]$) plus a DC offset.

$$\Omega_{xx}[n] = \delta'[n] - \frac{1}{L+1} \tag{15}$$

Rife & Vanderkooy (1989) also pointed out that measurements done by using the symmetrical MLS (thus formed by ± ones) are virtually AC coupled, meaning that the DC component is totally rejected. This feature of MLS to possess an autocorrelation equal to a unit-sample makes it a very valuable tool to measure impulse responses. The output of a system due to a (periodic) MLS input, will also be periodic, as well as its impulse response $h'[n]$, thus

$$y[n] = x'[n] \circledast h'[n] \tag{16}$$

where t $y'[n]$ is the periodic (modulo N), respectively. The cross-correlation between the input and output signals gives the (periodic) impulse response $h'[n]$ of the system, since

$$\Omega_{xy}[n] = \sum_{k=0}^{L-1} x'[n]y'[n-k] \; = x'[n] \circledast y'[n] \tag{17}$$

Then using Eq. (16) in Eq. (17) it results in

$$\begin{aligned}\Omega_{xy}[n] = {} & x'[n] \circledast (x'[n] * h'[n]) = (x'[n] \circledast x'[n]) * h'[n] \\ & = \Omega_{xx}[n] * h'[n]\end{aligned} \tag{18}$$

and then using Eq. (23) in Eq. (26), gives

$$\Omega_{xy}[n] = \left(\delta'[n] - \frac{1}{L+1}\right) * \mathrm{h'[n]} = \mathrm{h'}_{\mathrm{AC}}(n) - \frac{1}{L+1}\mathrm{h'}_{\mathrm{DC}}(n) \tag{19}$$

where $\mathrm{h'}_{\mathrm{AC}}(n)$ is actual periodic (ac coupled) periodic impulse response (PIR) of the system, and $\frac{1}{L+1}\mathrm{h'}_{\mathrm{DC}}(n)$ represents the residual DC offset (which is null for symmetric MLS usage). To fulfil the prerequisite of circular convolution, an MLS sequence must be used to excite the system under examination to a steady state, before actually performing measurements Rife & Vanderkooy (1989)(. Moreover, to avoid aliasing, due to the circular convolution, the MLS duration has to be equal to or longer than the RT of the room/system in question.

2.4.4.2 Fast Hadamard Transform

Albeit MLS was initially used within the frequency domain FFT is not applicable to handle MLS signals, since an additional bit must be used in order to obtain an even number of elements of the MLS sequence and this will transform the MLS into something slightly different. Subsequently, MLS deconvolution was found to be more efficiently performed using the fast Hadamard transform (FHT), directly in the time domain, as demonstrated by Alrutz & Schroeder (1983) and Chu (1990), and

proposed by Hee (1990). Similar to the FFT, the FHT requires $N \cdot \log_2 \mathrm{N}$ operations. The FHT is a 2^Nx 2^Nmatrix of ±1, scaled by a normalisation factor – typically $\frac{1}{\sqrt{2}}$. It can generate recursively starting from the unitary identity matrix $H_0 = 1$, and then define

$$H_N = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{N-1} & H_{N-1} \\ H_{N-1} & -H_{N-1} \end{pmatrix} \quad \text{for N>0} \tag{20}$$

The matrix transforms an arbitrary input vector of 2^N real numbers into 2^N real number output vector. The transformed data are then reordered and normalised.

2.4.4.3 Practical Aspects of MLS

The downsides and limitations of the use of MLS in acoustic measurements as highlighted in Dunn & Rife (1994), Dunn & Hawksford (1993), and Vanderkooy (1994), are chiefly inherent to its susceptibility to non-linearity and time-variance. They are summarised in the following.

Time Variance Susceptibility

For RIR analysis, the condition of time invariance, which is one of the assumptions behind LTI systems, is unlikely to be guaranteed for several practical reasons. Air drifts, temperature changes, and people's movements modify the physical states of enclosures. This results in RIRs that can be slightly different from measure to measure. The assumption of time invariance will therefore hold true only if it is slowly compared with the measurements duration. For example, the BS EN ISO 18233:2006 standard sets the amount of variance that can be accepted during measurements, and

also highlights that some measurement methods are more or less prone to incur in time-variance artefacts. For example, it states that MLS, and more generally noisy signals, are more susceptible to time variance than chirp signals.

As reported in Vorlander & Kob (1997), MLS based measurements can be affected by two types of time variance, “intraperiodic” and/or “interperiodic”. The intraperiodic relates to quick time variance effects occurring during a single MLS (i.e. air drift produced by the ventilation/air conditioned systems), and is particularly detrimental when long MLS sequences are used (e.g. when measuring large halls). On the other hand, interperiodic refers to slow time variance effects (like slow changes in temperature), which may affect the efficiency of a synchronous averaging.

Svensson & Nielsen (1999) estimated that intraperiodic and interperiodic time variance causes an “apparent” noise in the recovered RIR, of noise level by 6 dB /and 12 dB per octave respectively. Time variance affects in greater measure both higher frequency and the late part of the RIR, as it is more sensitive to small variations of time of flight path of acoustic waves, and as a consequence, the measured RT may be underestimated (Bradley, 1996).

Detecting environmental changes like temperature and airflow can also be used to monitor the degree of variance during measurements, or as suggested by Svensson & Nielsen (1999), post-averaging can help to control the amount of variance introduced in an average measurement by discarding the terms that exceeded a determinate threshold. It should be noted, however, that Satoh et al. (2002) reported that a

temperature change of 0.1 degrees already leads to gross errors in the MLS-based measurements, with the high frequencies being the most affected.

Noise Rejection and Averaging

An attractive feature of an MLS is its ability to deal with noise. Its pseudorandom nature phase-randomises every extraneous noise, both impulsive and steady, with the consequential effect of spreading the noises along the recovered RIR. Moreover, since MLS is also a deterministic signal it can be synchronously averaged leading to a +3 dB increment of the SNR for each doubling of the number of averages, if background noise is uncorrelated. This, however, holds only true when time variance is negligible, as repeated measurements will give exactly the same responses. The reason is because identical system outputs accumulate over a large number of repetitions, while the background noise is uncorrelated among these repetitions and will be averaged out Rife & Vanderkooy (1989). The length of MLS used for a measurement is commonly chosen 2 to 5 times longer than the (supposed) RT of the room (Stan et al., 2002).

The MLS noise immunity, and averaging techniques allow measurements to be performed under very low or even negative signal-to-noise ratios (Nielsen, 1996), which gives rise to the idea of using MLS for occupied measurements. However, it has been proven (Serafini et al., 2010) that time variance will eventually halt the SNR increment due to the averaging process, making it reach a maximum value before decreasing to lower values. Lengthy measurement sessions are therefore unlikely to produce useful results.

Satoh et al. (2002) demonstrated that MLS measurements using synchronous averaging could lead to gross errors in the estimation of RT from a temperature change of 0.1 degrees. Therefore, MLS averaging must be performed with extreme care, especially when measurements are performed in large spaces (i.e. an auditorium or a cathedral) where temperatures may change very quickly and air drifts may be quite substantial.

Artefacts and Crest Factor

Acoustic spaces like rooms, halls, etc. are perfect linear systems. Nonetheless, the devices used for acoustic measurements tend to show some degree of non-linearity especially when operating close to saturation levels, i.e. high amplitude/volume. Vanderkooy (1994) accurately explained the MLS sensitivity to non-linearity, and showed that peak-like artefacts will appear along the deconvolved RIR. He also noted that artefacts are related to the particular MLS used, and then published a list of sequences that had a better behaviour when applied to non-linear systems.

However, if MLS with different lengths or tap-configurations are averaged together artefacts will be reduced because artefacts appear in specific positions along the RIR, which is dependent on the particular MLS used (Greest & Hawksford, 1995). Commonly, a useful, but very time-consuming practice to reduce artefacts consists of carefully calibrating the amplifiers by selecting the maximum MLS level that produces the least amount of distortions. It is worth noting that, although the theoretical CF of an MSL is of 0 dB, the digital-analogic converter and the anti-aliasing filter of the measurement device can lead to a CF as high as 11 dB (Müller & Massarani, 2001).

Hybrid MLS

Paulo et al., (2009), proposed the use of a particular technique to increase the SNR in MLS measurements. Two approaches were used and combined to form what he called a hybrid MLS. A set of eight ML sequences (only a 15^{th} and a 17^{th} order MLS were tested), were fed into a room and the room's response acquired. To simulate sing such method in occupied conditions, disturbances like speech or music (various genres) were used during the measurements.

The first part of the method consists of dividing each sequence into "n bits" long slices (for n = 1024 for speech and n = 2048 for music, the method gives the best results, as stated in Paulo's work) and measuring the mean square energy (MSE) of each of them. Since the MLS power is constant in all sequences but disturbances vary, by comparing the MSE of each slice with their homologues in each sequence, a new ML sequence can be built using the slices with the lowest SNRs. In his work Paulo's described the procedure as a "noise scrambler", which gives an MLS with the lowest possible SNR.

The second part of Paulo's method involves a multi-band filtering of each sequence (both third-octave and octave bands filter bank were experimented), and similarly to the first method each sub-band is compared to their homologues to find the ones with the lowest MSE. However, he also pointed out that in the case scenario where all homologous slices have low SNR, the method gives worse results than when all sequences are averaged together.

Paulo's work concluded that the combination of both techniques (hybrid MLS) gives an overall improvement of the SNR, compared with the standard MLS technique, as high as 24 dB for speech disturbances and between 6 and 10 dB for music disturbances (depending on the music genre). In general, he stated that the method achieves a better SNR than conventional MLS techniques in presence of high non-stationary background noise; an advantage that may be exploited to reduce the measurement time in slightly time-variant systems.

Inverse Repeated Sequences (IRS)

To overcome some of the limitations of MLS, mainly its non-linearity susceptibility, a different version of the technique named 'inverse repeated sequence' (IRS) was proposed to partially solve the problem (Ream, 1970). The basic idea behind the IRS technique was to form an MLS sequence by joining two equal MLS sequences but with the second being an inverted copy of the first (i.e. inverting the signs of the MLS bits), as proposed by Dunn & Hawksford (1993). In such a way the odd-order distortions will be cancelled out, which is a real advantage only for those systems that exhibit such types of distortions. Another drawback is that the deconvolution of the system's response cannot be done using the FHT, but it requires the classic FFT based deconvolution (Angelo Farina & Righini, 1997). Moreover, as proved by Stan et al. (2002), differences between MLS and IRS tend to disappear when a low-level MLS is used because of the minor introduced distortions.

2.4.5 Dual Channel FFT

2.4.5.1 Theoretical Description of the Dual Channel FFT

The advent of fast Fourier transform (FFT) analysers brought new tools for acoustic and electroacoustic applications (Thrane, 1979; Bendat & Piersol, 1993). In particular, a new technique named the "dual channel FFT" showed interesting features that could be used for occupied acoustic measurements. The dual channel FFT technique makes it possible to measure the frequency response function of an unknown system, by dividing the Fourier transform of its two channels inputs, thus

$$\frac{Y(f)}{X(f)} = H(f) \tag{21}$$

Note that the signals must be a non-zero continuous function of frequency (in the measured range), and their Fourier transforms must exist. Commonly, the measurement of the FRF is obtained by using a modified version of Eq. 9, which can provide additional advantages. For example, as described in Herlufsen (1984a, and 1984b), two slightly different transfer functions can be calculated as the ratio of the cross-power spectrum densities (CPSDs) and the auto-power spectrum densities (APSDs), thus

$$H_1(f) = \frac{S_{xy}(f)}{S_{xx}(f)} \tag{22}$$

$$H_2(f) = \frac{S_{yy}(f)}{S_{yx}(f)} \tag{23}$$

where $S_{xy}(f)$ and $S_{yx}(f)$, and $S_{xx}(f)$ and $S_{yy}(f)$ are the cross-spectra and the auto-spectra of the input (x) and the output (y), respectively. One advantage of using $H_1(f)$ or $H_2(f)$ depends on "where" the noise is expected to contaminate the system in question. In practice, when a noise $n(t)$ mixes with the input signal $x(t)$, which will become $x_{noisy}(t) = x(t) + n(t)$, $H_1(f)$ will guarantee a higher SNR; conversely, if the noise will add to the output of the system $y(t)$, giving a noisy output $y_{noisy}(t) = y(t) + n(t)$, then $H_2(f)$ should be preferred. This latter case scenario is generally identifiable with acoustic measurements, since the background noise biases the measured output of the system. The former case scenario, or a combination of both, may be more useful in the case of acoustic measurements performed with the dual channel FFT, as background noises affect both the input and the output.

Impulse responses can be obtained from FRFs through FFTs, as they are transform pairs, as reported in Section 2.3.2. Others functions can be obtained by the complex relation between two channels, each generally used to analyse the relationship between the two channels under a different point of view, as pointed out in the scheme shown in Figure 5 (extracted from Bruel&Kayer technical review by H. Herlufsen, 1984).

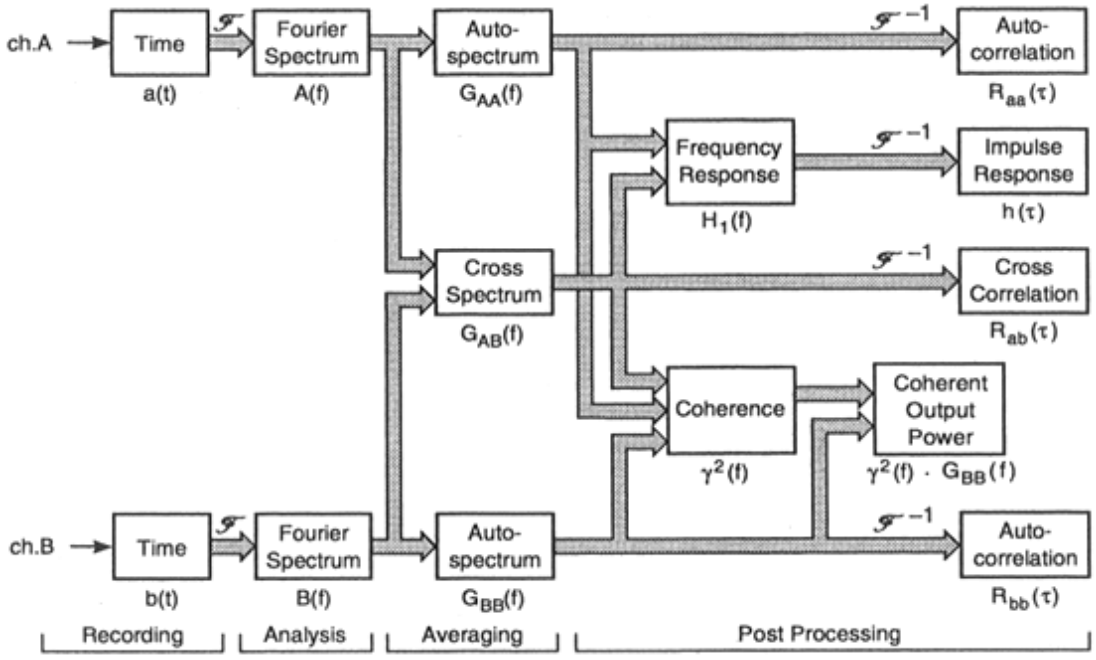


Figure 5 – Simplified block diagram of a dual channel analyser. Reported from Bruel&Kayer technical review, 1984.

For measuring the degree of correlation between input and output, a coherence function $\gamma^2(f)$ has been defined as

$$\gamma^2(f) = \frac{H_1(f)}{H_2(f)} = \frac{S_{xy}(f)S_{yx}(f)}{S_{xx}(f)S_{yy}(f)} \tag{24}$$

The coherence function $\gamma^2(f)$ expresses how much of the input signal is also present in the output of the system. The coherence function is bonded between $0 \leq \gamma^2(f) \leq 1$, where 0 and 1 stand for a completely uncorrelated and a completely correlated output in respect to the input. The coherence function is generally used to validate the accuracy of measurements, which should be considered valid only when $\gamma^2(f) > 0.9$.

It should be noted that the coherence function also quantifies the degree of linearity between input and output, thus coherence will decrease when the system exhibits a certain amount of non-linearity. However, in dual channel FFT analysis, the distortions introduced by the loudspeakers become part of the input signal, and therefore they will not affect the coherence function. When noise is both present at the input and output $H_1(f)$ and $H_2(f)$ represent the lower and upper limit within which the true transfer function $H(f)$ lies. In the absence of noise $H_1(f)$ and $H_2(f)$ are equal to each other, and also to $H(f)$.

2.4.5.2 Use of the Dual Channel FFT for live measurements

With the dual channel FFT technique, any broadband frequency-continuous excitation signals can be used as stimuli to measure the FRF of the system, and therefore even music. In fact, in 1992, Meyer experimented measuring the FRF of a concert venue using the playing band's music as the input of the two channels connected to an FFT analyser. However, the spectral distribution and amplitude level of live music is quite erratic and may have sparse frequency content and low amplitude at some frequencies. This makes averaging unavoidable to reduce the signals' variance and to increase the SNR. A sparse signal, as defined by Meyer himself, is a signal in which not all frequency components of the interested frequency interval are present during a measurement period. As a result, not all frequencies may be excited and/ or the SNR for some frequencies may be too low. Missing frequencies or near zero values may cause inaccurate or wrong FRF measurements.

In practice, averaging is always necessary to obtain a smoother measurement as well as to increase the SNR. This may lead to lengthy measurements, in order to achieve a

sufficient accuracy, which strongly depends on the signals' characteristics, and as pointed out by Schoukens et al. (1988), "*if a relative accuracy is demanded over the measurement band, the frequency with the minimal SNR will determine the measurement time*".

Nonetheless, over the years, algorithms that are more sophisticated have been developed, achieving better efficiency and accuracy in live measurements, which led Meyer to develop the first SIM (source independent measurement) audio analyser system, a measurement tool for sound engineers to be used for assessing loudspeakers' FRF during live concerts. Nowadays, the latest version of SIM is SIM 3 (2003), and other similar products, like the SMAART© (v.7) software and EASERA (v. 1.2), are also currently used in the live music industry. However, although these products are the "state-of-the-art" for public address systems (PA) alignment and testing, the obtained FRF still struggles to pose the same accuracy and reliability of classic measurement methods. However, it should be pointed out that it is the use of non-deterministic signals, rather than the technique itself, that determines the lower performances of the dual channel FFT (Müller & Massarani, 2001), compared with other measurements technique, which use instead deterministic signals.

Moreover, beside the uncertainties that occur due to source signals, another practical problem affects the reliability of the dual channel FFT technique for accurate RIR measurements. RIR are a source-receiver relationship, ideally both represented by infinitesimally small point in the space. As a physical source-point is clearly not realisable, an omnidirectional dodecahedral loudspeaker is used as the source point and an omnidirectional microphone as the receiver point (standard ISO 3382-2:2008).

This configuration, works for single channel RIR measurements. However, in this way is impossible to have distortion-free measurements.

A solution to allow using two channels is to place a microphone in the "virtual" central point of the live orchestra/band playing. In this way, one channel will be used to capture the source sound while the second channel will capture the propagating sound in a place in the room. However, this solution pretends that a wide sound source is assimilated to a point-wise source, which is clearly in contrast with one of the aforementioned standard requisites.

2.4.6 Sine Sweep Based Methods

The TDS technique discussed in paragraph 2.4.3 is based on the use of (linear) chirps, however, in Heyser's method (1967) the deconvolution is analogically implemented through the use of a tracking filter. Thanks to better performances of computers and improved digital system processing (DSP) techniques (such as the fast deconvolution technique proposed by Kirkeby et al. (1998)), nowadays deconvolution can be directly implemented in the digital domain. This paragraph examines two measurements methods, which employ frequency-varying chirps, but the deconvolution of the output of the system is digitally implemented and based on the use of inverse filtering. From a system's theory point of view, a time-varying signal can be represented in polar form as

$$x(t) = A \cdot \sin\left(\theta(t)\right) \tag{25}$$

where $\theta(t)$ is the varying phase. A linear chirp

$$f(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} = f_0 + kt \tag{26}$$

is a signal where the instantaneous frequency $f(t)$ varies linearly with time where f_0 is an initial frequency value and k defines the chirp-rate ($k = B/T$), i.e. the ratio of the signal bandwidth ($B = f_2 - f_1$) over the signal duration T. The phase can be calculated by integrating Eq. 17 over time

$$\theta(t) = 2\pi \int f(t)dt = 2\pi \int (f_0 + kt)dt = 2\pi \left(f_0 t + \frac{k}{2}t^2 + \phi_0\right) \tag{27}$$

using Eq. 12 in Eq. 9 and knowing that $k = \frac{f_2 - f_1}{T}$ and $\phi_0 = 0$; it results in

$$x(t) = A \cdot \sin\left(2\pi\left(f_{\text{in}}t + \left(\frac{f_{\text{fin}} - f_{\text{in}}}{2T}\right)t^2\right)\right) \tag{28}$$

where A is the amplitude; T the chirp duration; f_{fin} and f_{in} are the final and the initial frequencies, respectively. In the digital domain it becomes

$$x[n] = A \cdot \sin\left[2\pi\left(f_{\text{in}}n + \left(\frac{f_{\text{fin}} - f_{\text{in}}}{2N/f_s}\right)n^2\right)\right] \tag{29}$$

where N is the total number of samples (n is the samples' index) of the sequence and f_s is the sampling frequency.

2.4.6.1 Inverse Filter Deconvolution

The response of the system $y(t)$ to an excitation signal $x(t)$ is described by the (linear) convolution of the signal with the RIR of the system $h(t)$, thus $y(t) = x(t) * h(t)$. To retrieve the response of the system, a cross-correlation technique can be used by deriving an appropriate inverse filter to "erase" the presence of the excitation signal in the measured response of the system (J. Bendat & Piersol, 2011). In practice, an inverse filter $r(t) = x(-t)$ is constructed by time reversing the excitation signal. Then, convolving $x(t)$ with $r(t)$ gives the delayed Dirac delta function $\delta\,(t-\tau)$, thus $x(t) * r(t) = \delta\,(t-\tau)$. The time lag τ corresponds to the length of the signal $x(t)$, and a simple time shifting can be used to remove the delay and set the start of the recovery RIR a t = 0. The RIR is obtained by convolving the system output $y(t)$ with the inverse filter $r(t)$, thus $y(t) * r(t) = \big(x(t) * h(t)\big) * r(t) = \delta\,(t-\tau) * h(t)$, and by removing the delay τ yields the sought $h(t)$.

2.4.6.2 Exponential Sine Sweep Technique

While experimenting with solutions to the limitation of MLS to cope with slightly non-linear systems, Farina (2001) proposed the use of an exponential sine sweep (ESS) technique, a chirp whose instantaneous frequency varies exponentially. An ESS is characterised by a pink-shaped PSD (-3 dB/octave), which allows more energy to be fed at lower, rather than at higher, frequencies. This compensates for the (typically) higher background noise at low frequency bands. The energy fall is compensated while deconvolving the response of the system by applying a +6 dB/octave amplitude envelope to the deconvolution filter (matched filter). An important advantage of using an ESS signal is that harmonic distortions, such as those caused by loudspeakers,

since they have their instantaneous frequency unrelated with the used matched filter will be repacked in specific time lags, before the casual part of the RIR. This permits of calculating, in a single measure, the linear response and each order of distortion. Using long sweeps increases the distance between harmonics, and avoids them overlapping (due to the reverberant tail) and facilitates their extraction by time windowing. An ESS is mathematically defined as

$$x(t) = \sin\left(\frac{\omega_1 T}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\left(e^{\frac{\ln\left(\frac{\omega_2}{\omega_1}\right)t}{T}} - 1\right)\right) \tag{30}$$

where T is the length of the signal in seconds and $\omega_1 = 2\pi f_1$ and $\omega_2 = 2\pi f_2$ are the starting and the ending angular frequency of the sweep respectively.

2.4.6.3 Linear Sine Sweep Technique

A linear sine sweep (LSS) is constructed, letting its instantaneous frequency vary linearly with time. LSS presents some differences compared to the ESS. Firstly, the LSS spectrum is white, therefore constant to all the frequency range; secondly, the distortion artefacts will appear superimposed before the IR and hence cannot be analysed separately. On the other hand, the LSS is still able to retain only the linear part of the RIR (Müller & Massarani, 2001). An LSS signal can be defined as:

$$x(t) = \sin\left(\omega_1 t + \frac{(\omega_2 - \omega_1)}{2T} t^2\right) \tag{31}$$

where the sweep ranges from the starting angular frequency ω_1 to the final value ω_2 in a time T. It should be noted that the use of linear chirps, apart from the TDS technique, has also been used for acoustic measurements by some Japanese researchers (Aoshima, 1981; Suzuki et al., 1995), who preferred to call the technique the time stretched pulse (TSP) technique.

2.4.7 Pros and Cons of Chirps

Chirp-based measurements have been proven to yield superior results compared with other measurement techniques (Müller & Massarani, 2001; Stan et al., 2002), and when measurements are only aimed at measuring the linear part of RIRs, using exponential or linear chirps would then give the same results (Farina, 2001). In fact, an important advantage of using chirp-based methods is related to the possibility of discarding the non-linear harmonic distortions introduced by the measurement chain from the RIR, leaving only the linear part of the system transfer function. This occurs because a linear deconvolution of the response of the system by a matched filter will push the distortions to appear on the time axis before the RIR. Distortions can then be simply rejected by (time) gating them out. A practical consequence of this distortion rejection property is that more power can be used for the measurements. This leads to a much higher SNR compared to other measurement methods (Stan et al., 2002). It has also been demonstrated that chirps are less prone to incur in time variance induced artefacts, as frequencies are excited once at time (Müller & Massarani, 2001); although synchronous averaging still poses a threat, especially at high frequencies (Farina, 2007). To circumvent time variance problems, a single long chirp is preferable to avoid or limit averaging, as also suggested in the BS EN ISO

18233:2006 standard. Another notable benefit of using chirps is the possibility of using signals shorter than the rooms' RT (conversely than for MLS, for example). The linear convolution actually removes the risk of time aliasing. Although the late part of the RIR could be lost, in practical measurements a silenced part is added to the end of the chirp to allow the measurement to capture the whole RIR.

On the other hand, chirp based measurements may incur in artefacts provoked by the transient start and stop of the excitation signal. Such sharp transients cause amplitude fluctuations at the edge frequencies of the chirps' spectra, which will also be increased if step discontinuities are generated by a non-zero crossing truncation of the chirp signal. Those amplitude fluctuations before and after the main lobe are called pre-ringing and post-ringing, respectively. Pre-ringing is acoustically more detrimental for the audio quality as it can be heard easily. On the contrary, the post-ringing results almost unnoticeably due to the masking effect. For exponential sine sweep based measurements, the tendency is to remove only the pre-ringing and, as suggested by Farina (2000), this can be achieved by applying an amplitude fade-in to the time signal. However, as the measurements are performed to almost the whole audible spectrum, the measurement system itself (mainly the amplifiers and loudspeakers) induces an amplitude fade-out to the signal when the frequency is close to the system Nyquist's point. This implies an inherently applied fade-out to the time signal. Farina (2007), also suggested using chirps that have broader spectra than the one that is being measured, as artefacts will be pushed out over the interested frequency band. Instead, Novák et al. (2009) proposed to phase-align the chirps to avoid discontinuities at the signals' extremities, however, even in doing so, overshoots and undershoots can still be as high as ± 3 dB (Vetter & Di Rosario, 2011).

2.4.8 Using Chirps for Occupied Measurements

The loud and high-pitched chirping sound causes annoyance to the audience much more than the noise-like sound of MLS, or live music using the dual channel FFT. This has so far hindered using chirps for occupied measurements. Decreasing the level of chirps reduces loudness but also decreases the SNR. In contrast with MLS, chirps do not have a phase randomisation property, which hinder their use with low or negative SNR, and makes them more susceptible to impulsive noises, which may occur during measurements. Such susceptibility could be enhanced further when the measurements are taken in occupied spaces. Impulsive noises happening during the measurements create an "inverse chirps" in the deconvolved impulse response, which will then invalidating the measures. As suggested by Farina (2007), the impulsive noise could be manually erased by digitally removed the affected portions of audio in which the noise event has occurred, This actually corresponds to notch filtering the RIR spectrum at frequencies that depend on the erased time intervals.

Despite these drawbacks, some measurements have been performed using chirps. For example, Beranek (2004) described some measurements carried out in occupied halls by employing two linear chirps; one used to excite the low frequency band up to 500 Hz, the other used to excite mid and mid-high frequency bands. This is performed in order to achieve a higher SNR at a low frequency where background noise is higher. The audience voluntarily took part in those measurements. Recently, Satoh et al. (2013) proposed a method in which broadband exponential, or linear chirps were inserted into the music with the intent of reducing chirps' obtrusiveness. The chirps were both inserted between phrases and with the music. A third case, which used a slow varying chirp overlaid with the music, was also experimented. The impulse

responses obtained in all the scenarios showed spurious components in the squared IRs' decay curves due to the presence in the music. However, for the first case, when chirps were separated from the music, results were more promising. The authors concluded that further studies were necessary to improve the measurements' SNR and accuracy. Details about the actual limitation of chirps' obtrusiveness have not been given, and further research has not been carried out so far.

2.5 Blind and Semi-blind Estimation of Acoustic Parameters

During the years, methods have been developed with the aim of extracting acoustic parameters directly from the reverberant sound energy decays, without (or almost) a priori knowledge of the original source. Such methods can be broadly grouped into semi-blind and blind deconvolution techniques.

2.5.1 Overview

The idea of deriving RT from sound decays has been experimented with since the early eighties (Cremer et. al., 1982). The RT and the EDT were calculated from the sound decay consequential to the sudden stop of music played by an orchestra. This approach is generally known as the stop-chords method. Hidaka et al. (1998) carried out several measurements utilising this technique. Their research aimed at addressing the relationship between occupied and unoccupied measurements, and they developed an equation to relate acoustic parameters measured in unoccupied conditions with their supposed values in occupied conditions. The main drawback of this solution is the temporal smearing of the "end of music" caused by the intrinsic reverberation of musical instruments, and the impossibility to obtain a precise and sharp stop of all musicians. In addition, frequency fullness may remain unfulfilled for certain frequency bands, which may result in a false estimation of the RT. However, the idea of analysing acoustic parameters in such a way has been the precursor to more reliable and accurate procedures. Accurate analyses are nowadays performed employing innovative DSP algorithms, as briefly discussed in the following Sections.

2.5.2 Semi-blind Methods

Semi-blind methods are based on a partial knowledge of the probe signals. Algorithms are used to find the transient parts of the incoming sound, where the RT may be estimated using a linear-fit regression line and Schroeder's backward integration (Vesa & Härmä, 2005). Since the long pauses and transient nature of speech makes the presence and discovery of energy decay with large dynamic range more likely, Cox et al. (2001) and Li et al. (2001), proposed the use of artificial neuronal networks (ANNs), trained with known RIR, to extract reverberation times from speech utterances. Nonetheless, the method showed great accuracy, a long training phase of the ANN was required and validations were limited to software simulations. Moreover, the method lacked validations in real rooms and was limited to being used with speech. In an attempt to overcome such limitations, Zhang et al. (2006), and Kendrick et al. (2007; 2008) extended the ANN functionality to also work with music. Nonetheless, despite an excellent accuracy in RT estimation and many validations through simulated acoustic responses, Kendrick stated that the method could be inaccurate when used in complex and irregular geometrical rooms' shapes, or when not enough training was given to the ANN. Moreover, measurements performed in environments with low RT achieve better accuracy.

2.5.3 Blind Methods

Another approach to derive acoustic parameters from unknown signals is with blind deconvolution techniques. Such methods work in situations where there is little or no knowledge of the probe signal. A method developed (Couvreur et al., 2001), has been proven useful for application as automatic speech recognition (ASR), and also

Ratnam et al. (2003; 2004) highlighted the possibility of using a maximum likelihood estimation (MLE) on the exponential energy decay of speech sounds to derive acoustic parameters. Successively, Li & Cox (2003; 2007) further developed the method making it able to accurately extract the speech transmission index (STI). Zhang et al. (2006) used blind separation to extract acoustic parameters, although the method only works for minimum phase systems, which is never the case of RIRs. Although promising, the blind deconvolution methods use a single exponential decay model to represent the sound decay, which does not mirror reality and limits the estimations accuracy.

2.6 Conclusions

During the years, several acoustic measurement methods have been developed with the aim of obtaining accurate and reliable measures of RIRs or acoustical parameters. Each method had been demonstrated to be more or less efficient and accurate on performing such tasks, depending on the measurement conditions, i.e. outside/inside measures, background noise, time variance, etc. The knowledge of methods' features aids the decision of which method may be more appropriate for a particular case scenario. However, measurements method works fine for unoccupied venues but still lack of usability for occupied location.

It has been pointed out that audience strongly alters the acoustics of the spaces. An audience also increases background noise and time variance, and thus acoustic measurements in occupied conditions become even more difficult. Ultimately, usually an audience is reluctant to being subjected to unpleasant signals for research purposes, particularly when taking part in unpaid studies.

Deriving FRFs from live music through the dual channel FFT technique, although straightforward for occupied measurements lacks accuracy due to the sparse frequency content of naturalistic music.

Since measuring the RIR of occupied acoustical spaces was the motivation of the Ph.D. project, solutions have been researched examining solely the methods that could guarantee the measure of RIRs, i.e. ESS, LSS, and MLS. In the following Section, the key points and limitations of MLS and chirp-based methods are detailed and listed. The symbol "plus" indicates an advantage, the symbol "minus" indicates a

disadvantage and the symbol "plus & minus" indicates neither an advantage nor a disadvantage:

MLS

+ Noise rejection due to phase randomisation
+ Autocorrelation function very close to an ideal Dirac delta function
+ Theoretical crest factor of 0 dB
+ Easy and cheap to be implemented
+ Well tolerated by audience at medium or low volume
+ Virtually effective even at negative value of the SNR when used in conjunction with an averaging technique, and in time-invariant systems
± The hybrid MLS (by Paulo et al. (2009)) gives a better SNR (than standard MLS) in the presence of non-uniform disturbances, although a very low SNR does not introduce evident advantages
± Non-linearity systems cause peaky artefacts along the RIR, although they could be partially eliminated by time gating the RIR
± Only odd order distortion can be totally avoided (when IRS is used)
− Measurements chain may increase the CF up to 11 dB
− Strong time variance susceptibility especially at high frequencies and when over-averaged

Chirps

+ Non-linearity rejection
+ Highest SNR among all methods
+ A single slow and long chirp can be used in place of many averages

+ Robust against slight time variance
+ Customisable spectra
+ Chirp length can be shorter than the RT, and providing a silenced part it guarantees the measure of the whole RIR
– Cannot be used for measurements with negative or very low SNR, e.g. inaudible measurements
– Impulsive noises produce residuals in the deconvolved RIR, which may invalidate the measures
– The high-pitched sound of chirps are far more obtrusive that the warmer noise-like sound of MLS

2.7 Proposed Hybrid Method

Two solutions have been developed to measure RIRs in two complementary frequency ranges. They address audience nuisance by exploiting the advantages of chirps and MLS signals, and by taking into account the use of musical or masked stimuli. The combined use of the two measured RIRs forms the "Hybrid method" presented in this thesis.

3 The Presto-chirps Method

The presto-chirps method is a novel measurement method based on the use of a set of band-limited linear chirps, thus the hence "presto-chirps", to analyse room impulse responses. The prefix "presto", which in music terminology refers to a movement that is executed "very fast", was introduced to mark the intention of using the presto-chirps as quick chirps. In fact, the method is specifically designed to measure the acoustical responses of occupied venues, i.e. auditoria, halls, rooms, etc. (Serafini & Li, 2011b; Serafini, 2015). The core idea behind the proposed method is to use the presto-chirps to "compose" musical sounding stimuli, and then use them to perform acoustic measurements. This solution would avoid causing nuisance to audiences during acoustic measurements, and therefore allowing occupied measurement to be carried out more routinely.

To facilitate the correlation between presto-chirps and music, the start and stop frequencies of presto-chirps are set equal to the frequencies of the musical notes (equal tempered scale) of an 88 key grand piano. Since the fundamental frequencies of a grand piano range from 27.5 Hz to 4.435 kHz, an 89^{th} note has been added to the set of 88 presto-chirps with the purpose of allowing the measurement of the third/octave band centred at 4 KHz, which has a higher cut-off frequency of 4.49 kHz (BS EN 61260:1996). Thus, the presto-chirps method can be used to measure the frequency range from 27.5 Hz up to 4698.6 Hz (for simplicity this range will hereafter indicated as 27 Hz – 4.7 kHz).

Presto-chirps have bandwidths equal to semitones, which increase as the frequency increases, and vary from 1.6 Hz to 263.7 Hz. The linear frequency varying nature of presto-chirps makes them similar to, in musical terminology, “glissando” notes.

The advantage of using glissando notes (presto-chirps) is that they excite every frequency of the measured range. This condition is necessary to achieve complete measurements of impulse responses in the specified spectrum. Chirps are known to allow for the separation of the non-linear components of the electroacoustic devices used in the measurement process, which enables the removal of non-linear artefacts. Chirps are also more robust against time variance, compared with other probe signals, which is a condition likely to be dealt with in occupied venues, due to the presence of HAVC systems, and temperature fluctuation.

In addition, the frequency and time selectivity of presto-chirps could help mitigate rustling and object falling noises, again likely to occur in occupied venues. Furthermore, the degree of flexibility in composing measurements’ probes allows specifically composed and tailored stimuli to achieve higher SNR in noisy environments. For example, the amplitude of presto-chirps could be adapted and varied Could be adapted and varied according to the background noise to better deal with temporary noisy sub-bands.

3.1 The Presto-chirps

Presto-chirps are linear chirps whose instantaneous frequency ranges between the frequencies of two consecutive musical notes. Mathematically, this can be written as

$$x_i[n] = A \cdot \sin\left[2\pi\left(f_i n + \left(\frac{f_{i+1} - f_i}{2N/f_s}\right)n^2\right)\right] \quad (32)$$

where $f_i = 2^{(i-69)/12} \cdot 440 \ \ [Hz]$, A is the amplitude, and N the number of samples. The integer, *i*, is used to indicate a specific presto-chirp, and it corresponds to the integer numbers used in the MIDI tuning standard. Table 2 shows the correspondence between the index *i*, notes, and frequencies.

i	*Note*	*Hz*	*i*	*Note*	*Hz*	*i*	*Note*	*Hz*
21	A0	*27.5*	**51**	D#	*155.6*	**81**	A5	*880.0*
22	A#	*29.1*	**52**	E	*164.8*	**82**	A#	*932.3*
23	B	*30.9*	**53**	F	*174.6*	**83**	B	*987.8*
24	C1	*32.7*	**54**	F#	*185.0*	**84**	C6	*1046.5*
25	C#	*34.6*	**55**	G	*196.0*	**85**	C#	*1108.7*
26	D	*36.7*	**56**	G#	*207.7*	**86**	D	*1174.7*
27	D#	*38.9*	**57**	A3	*220.0*	**87**	D#	*1244.5*
28	E	*41.2*	**58**	A#	*233.1*	**88**	E	*1318.5*
29	F	*43.7*	**59**	B	*246.9*	**89**	F	*1396.9*
30	F#	*46.2*	**60**	C4	*261.6*	**90**	F#	*1480.0*
31	G	*49.0*	**61**	C#	*277.2*	**91**	G	*1568.0*
32	G#	*51.9*	**62**	D	*293.7*	**92**	G#	*1661.2*
33	A1	*55.0*	**63**	D#	*311.1*	**93**	A6	*1760.0*
34	A#	*58.3*	**64**	E	*329.6*	**94**	A#	*1864.7*
35	B	*61.7*	**65**	F	*349.2*	**95**	B	*1975.5*
36	C2	*65.4*	**66**	F#	*370.0*	**97**	C7	*2217.5*
37	C#	*69.3*	**67**	G	*392.0*	**98**	C#	*2349.3*
38	D	*73.4*	**68**	G#	*415.3*	**99**	D	*2489.0*
39	D#	*77.8*	**69**	A4	*440.0*	**100**	D#	*2637.0*
40	E	*82.4*	**70**	A#	*466.2*	**101**	E	*2793.8*
41	F	*87.3*	**71**	B	*493.9*	**102**	F	*2960.0*
42	F#	*92.5*	**72**	C5	*523.3*	**103**	F#	*3136.0*
43	G	*98.0*	**73**	C#	*554.4*	**104**	G	*3322.4*
44	G#	*103.8*	**74**	D	*587.3*	**105**	G#	*3520.0*
45	A2	*110.0*	**75**	D#	*622.3*	**106**	A7	*3729.3*
46	A#	*116.5*	**76**	E	*659.3*	**107**	A#	*3951.1*
47	B	*123.5*	**77**	F	*698.5*	**108**	B	*4186.0*
48	C3	*130.8*	**78**	F#	*740.0*	**109**	C8	*4434.9*
49	C#	*138.6*	**79**	G	*784.0*	**110**	C#	*4698.6*
50	D	*146.8*	**80**	G#	*830.6*			

Table 2- Midi numbers (i), name of note and frequencies of the keys (plus one) of a grand piano. The index *i* varies from 21, note A0 (27.5 Hz), up to 110, note C#8 (4698.6 Hz).

The Linear chirp constructed using Eq.(32) presents a sharp transition between zero and their maximum value (A), which gives rise to high magnitude ripples in the frequency response of presto-chirps. This effect is shown in Figure 6, which presents the frequency response of the presto-chirps $x_{69}[n]$ (half second long).

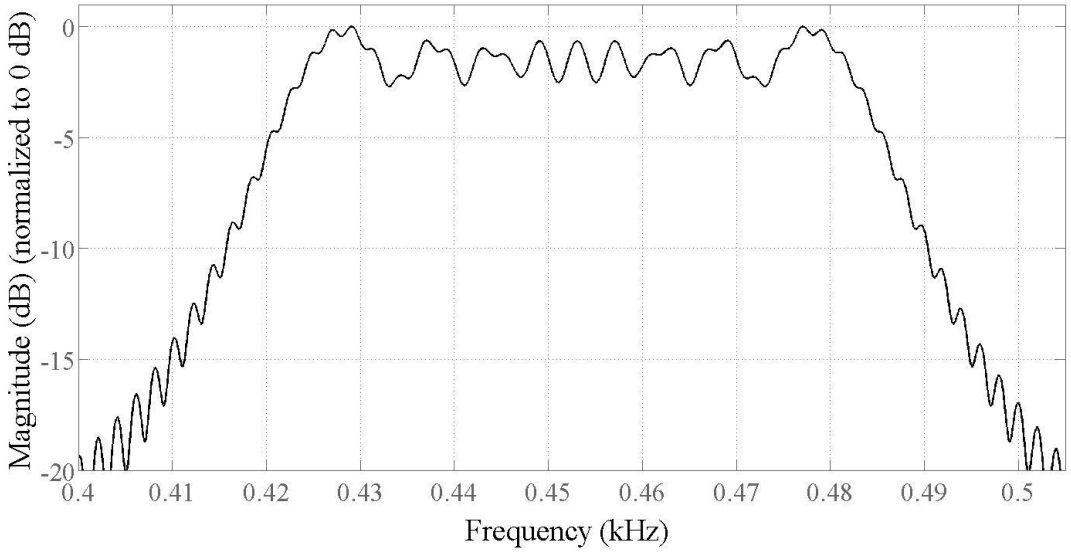


Figure 6 – Presto-chirp $x_{69}[n]$ – half second long.

Frequency ripples are the consequence of the impossibility of approximating band-limited signals through a finite sum of elements. In fact, an ideal band-limited signal has an infinite impulse response, but since only finite signals can be represented in real applications, signals are time gated to make them finite. Time gating can be seen in the frequency domain, as the convolution of a signal by a sinc function (which is the Fourier transforming of a rectangular window). Such convolution generates the ripples effect in the spectra of presto-chirps, as shown in Figure 6. Using "smoother" windows, rather than rectangular ones, can diminish ripples.

3.1.1 Tukey Windowing

Farina (2000) suggests applying a fade-in and, discretionally, also a fade out, which practically corresponds to a trapezoidal window to decrease the ripples at the edges of the spectra of exponential chirps. Starting from this suggestion, the presented research looked into the possibility of using Tukey windows, a.k.a. tapered cosine window (Harris, 1976, 1978; Nuttall, 1981) to fade in/out the presto-chirps in order to alleviate/remove ripples effects in the frequency responses of presto-chirps. Tukey windows are mathematically defined as

$$w_\alpha[n] = \begin{cases} \frac{1}{2}\left(1 - cos\left(\frac{2\pi}{\alpha N} n\right)\right), & 0 \leq n < \alpha\frac{N}{2} \\ 1, & \alpha\frac{N}{2} \leq n \leq N - \alpha\frac{N}{2} \\ \frac{1}{2}\left(1 - cos\left(\frac{2\pi}{\alpha N}(1-n)\right)\right), & N - \alpha\frac{N}{2} < n \leq N \end{cases} \quad (33)$$

where N represents the number of samples, and $\alpha \in [0,1]$. The parameter α defines the shape of the Tukey window, which becomes a rectangular window for $\alpha = 0$ and a Hanning window for $\alpha = 1$.

Values between zero and one lead to shapes varying from the two, as shown in Figure 7 for $\alpha = 0, 0.25, 0.5, 0.75, \text{and } 1$, which are the only values used in this thesis. Note that Tukey windows affect $\alpha\frac{N}{2}$ samples for side.

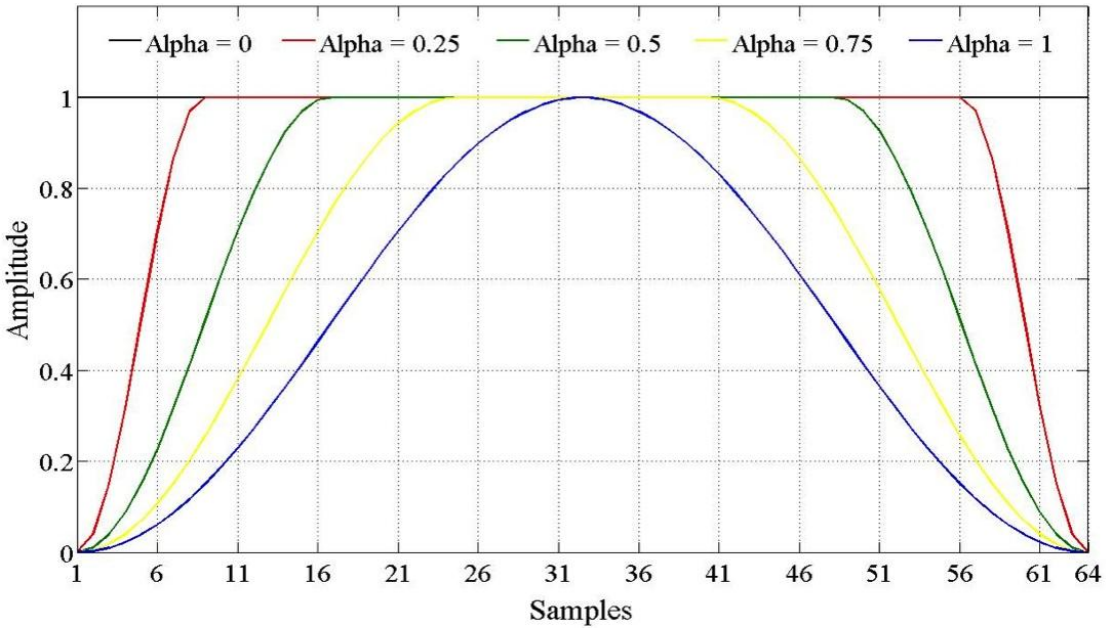


Figure 7 – Tukey windows' shapes for $\alpha = 0, 0.25, 0.5, 0.75,$ and 1.

The frequency responses of Tukey windows are reported in Figure 8 for the labelled α values. It can be seen that as α increases the magnitude of the side lobes decreases (leading to steeper roll-offs), and the width of the main lobe increases.

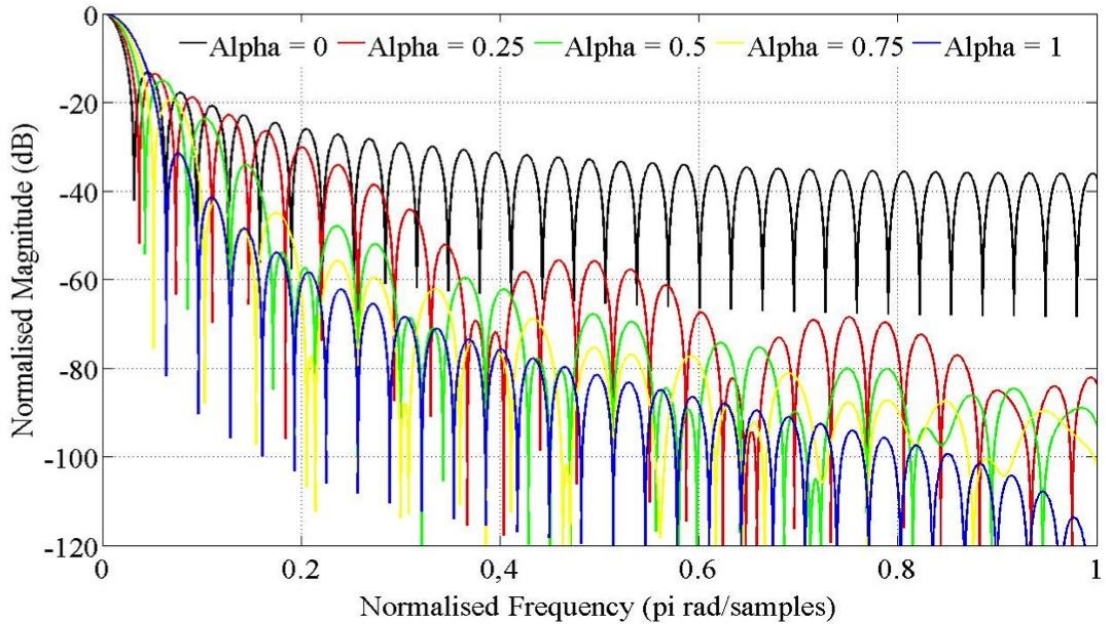


Figure 8 – Side lobes and roll-offs of for different Tukey windows.

Tukey windowed presto-chirps can mathematically be described as

$$\hat{x}_{i,\alpha}[n] = x_i[n] \cdot w_\alpha[n] \tag{34}$$

where $w_\alpha[n]$ is the used Tukey window, expressed as a function of α, and $\hat{x}_{i,\alpha}[n]$ is the resulting windowed presto-chirp $x_i[n]$. In the frequency domain, Equation (34) becomes

$$\hat{X}_{i,\alpha}[f] = \mathrm{X}_i[f] * W_\alpha[f] \tag{35}$$

where $\hat{X}_{i,\alpha}[f]$, $\mathrm{X}_i[f]$, and $W_\alpha[f]$ are the Fourier transforms of $\hat{x}_{i,\alpha}[n]$, $x_i[n]$, and $w_\alpha[n]$ respectively, and $*$ is the convolution operator. Figure 9 shows the spectrum of Tukey windowed presto-chirps varies as a function of α, as shown in where it reports the spectra of the presto-chirps $\hat{X}_{69,\alpha}[f]$ and $\hat{X}_{70,\alpha}[f]$ for $\alpha = 0.25, 0.5, 0.75, \text{and } 1$.

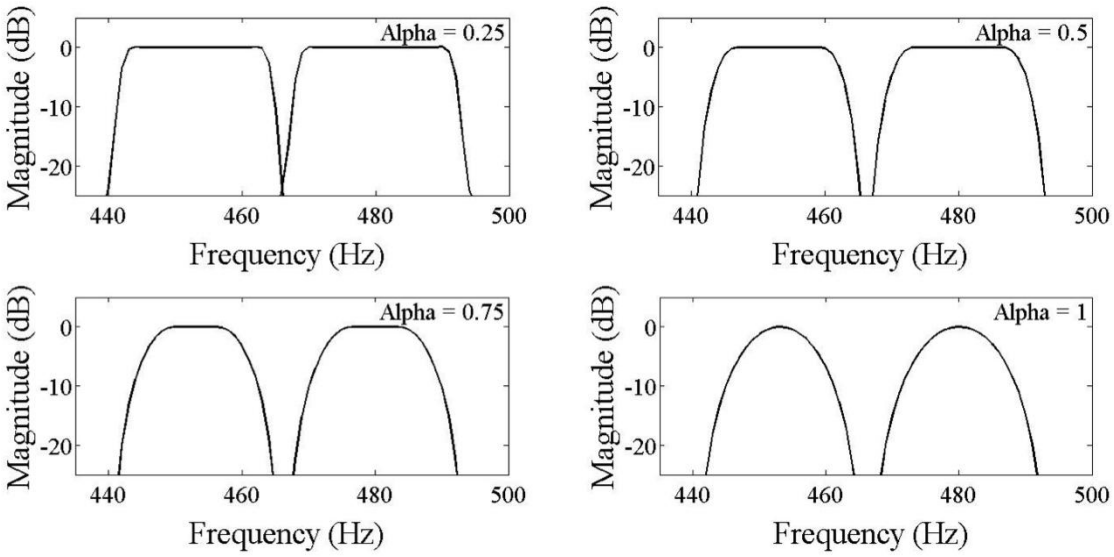


Figure 9 – Resulted spectra of two adjacent presto-chirps (Hz 440 – 466 Hz and 466 Hz – 494 Hz) for the labelled α values.

3.1.2 Mitigation of Frequency Discontinuities

Figure 9 reports frequency-adjacent Tukey-windowed presto-chirps. It can be seen that they present frequency discontinuities, which increases as α increases, due to the narrowing of the spectra of presto-chirps. To compensate for such frequency separations, since Tukey windows only affect the magnitude of those frequencies falling in the range from $\left[f_i, \alpha\frac{B}{2}\right]$ and $\left[\alpha\frac{B}{2}, f_{i+1}\right]$, the start and stop frequencies of presto-chirps are shifted as follows,

$$\begin{cases} f_{\mathrm{i},\alpha} = f_{\mathrm{i}} - \left(\frac{\alpha}{\beta}\right)\frac{B}{2} \\ f_{\mathrm{i+1},\alpha} = f_{\mathrm{i+1}} + \left(\frac{\alpha}{\beta}\right)\frac{B}{2} \end{cases} \tag{36}$$

where B = $f_{i+1} - f_i$ is the bandwidth. The parameter β has been calculated by trial and error, and it is used to allow the spectra of adjacent presto-chirps to intersect at -3.01 dB (which generally identifies the cut off frequency of filters), at frequencies equal to the frequency of musical notes. The value to be used of the parameter β varies depending on to the value of the parameter α, as reported in Table 3.

$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
$\beta = 1.35$	$\beta = 1.1$	$\beta = 0.85$	$\beta = 0.6$

Table 3 – Used β values based on the values of α.

Practically, it realises the situation outlined in Figure 10, where in dashed lines are represented as the spectra of the presto-chirps $\hat{X}_{69,1}[f]$ and $\hat{X}_{70,1}[f]$ prior to the

frequencies shift, and the solid lines represent them after it. Note how the -3dB magnitude intercepts the frequencies of the musical notes 440, 466, and 494 Hz.

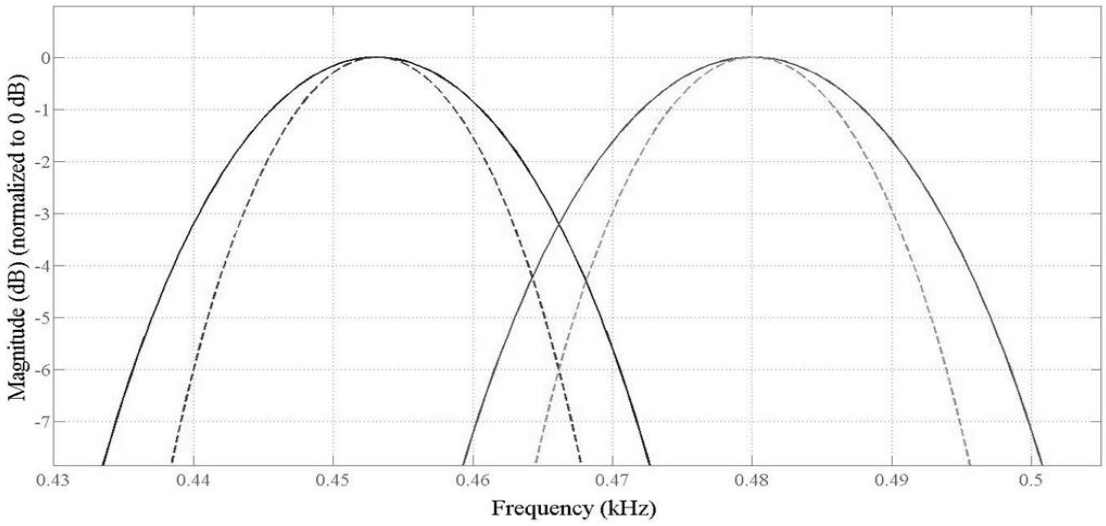


Figure 10 - Spectra of two adjacent presto-chirps ($\widehat{X}_{69,1}[f] = 440 - 466$ Hz and $\widehat{X}_{70,1}[f] = 466 - 494$ Hz). The dashed lines represent the presto-chirps prior the frequencies shift; the solid lines represent them after the shifting.

Figure 11 highlights that the frequency adjustment proposed in Eq.(36) works for all the used windows configurations, as can be seen in where the normalised magnitude responses, presented in linear scale for better representation of the phenomenon, of $\hat{X}_{109,1}[f]$ for $\alpha = 0.25, 0.5, 0.75, \text{and } 1$, are shown superimposed. Note that $\log_{10} 0.5 = -3.01$ dB is in a logarithmic scale.

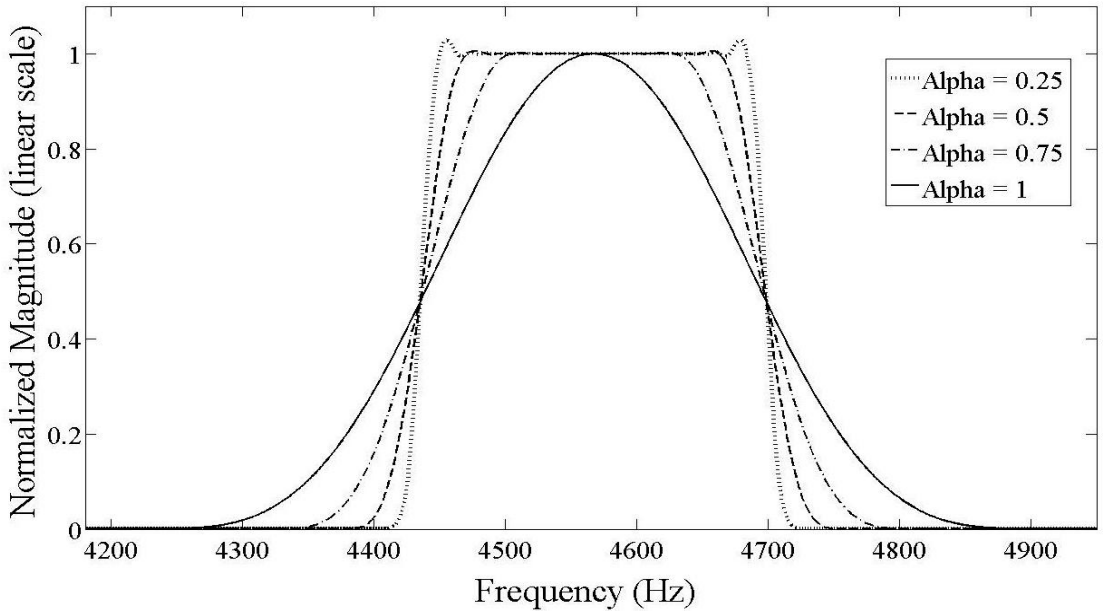


Figure 11 – Frequency response of the presto-chirp $\widehat{X}_{109,\alpha}[f]$ as a function of α.

The windowing and frequency shifts produce wider transition bands of the presto-chirps (which, using an analogy with filters, are the frequencies from below the half magnitude point to zero). Having presto-chirps with wider transition bands leads to progressively wider overlapped "areas" (the widening increases as α increases). Figure 12 shows this effect by plotting superimposed spectra of three frequency adjacent presto-chirps (each panel account for a specific α configuration).

The frequency overlapping does not cause errors per se, but it poses limitations on how stimuli can be composed, as is further discussed in Section 3.6.

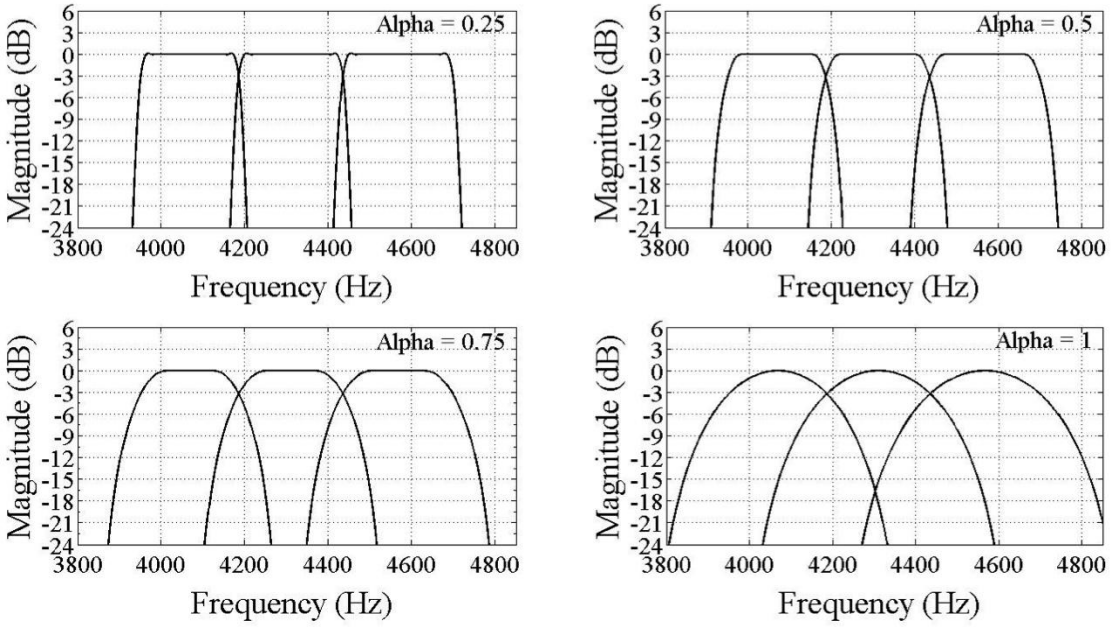
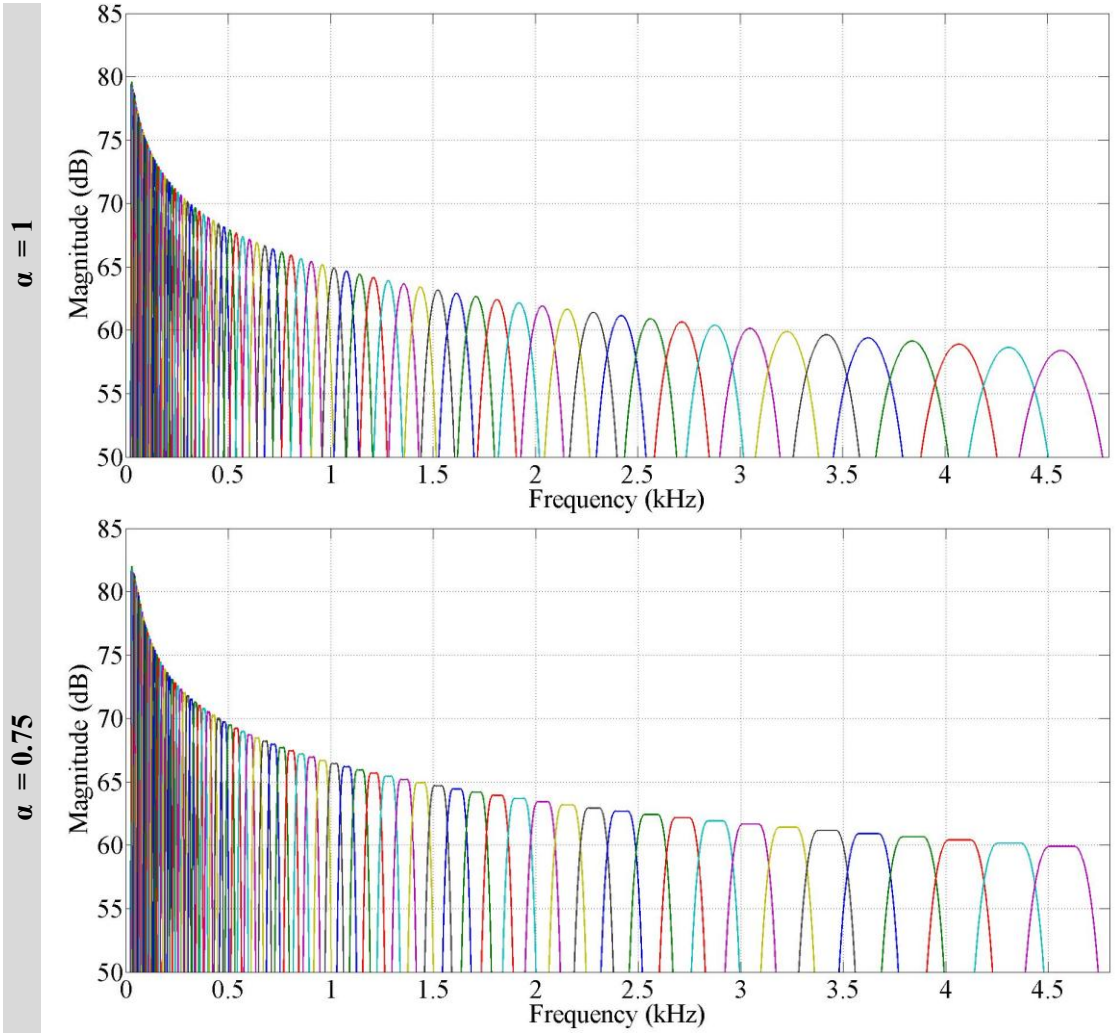


Figure 12 – Overlapping effect of three consecutive presto-chirps for different α values.

3.2 The Measurement Scheme

The 89 presto-chirps can be thought of as a filter bank that divides the audio spectrum from 27.5 Hz to 4698.6 Hz into logarithmically spaced sub bands. Each presto-chirp is used to measure a specific portion of such a spectrum. Figure 13 shows the spectra of the overall set of presto-chirps for Tukey windows equals to $\alpha = 1, 0.75, 0.5, \text{and } 0.25$, as labelled on the left.



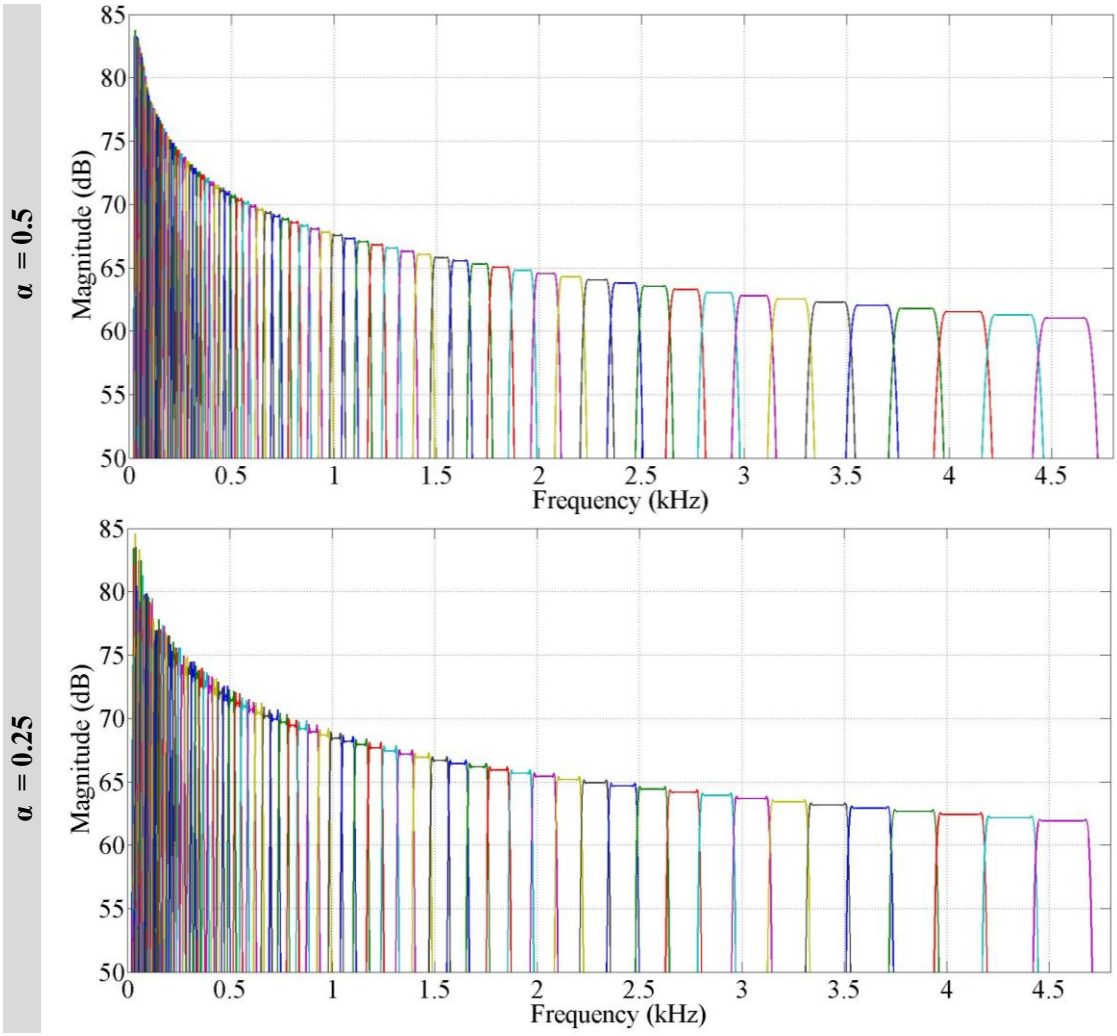


Figure 13 - Spectra of the overall set of presto-chirps for different Tukey windows.

The measurement scheme consists on the following steps:

- Creation of a stimulus using all the presto-chirps of a set at least once
- Feed the stimulus into the room under test
- Acquire and store the room's response
- Slice the recorded response into chunks
- Deconvolve each chunk by specific inverse filters to obtain a set of RIRs
- Weight and sum of all the partial impulse responses to form a broadband RIR

Figure 14 shows a schematic representation of the measurement scheme, from the partitioned audio, to the final sum of the partial impulse responses.

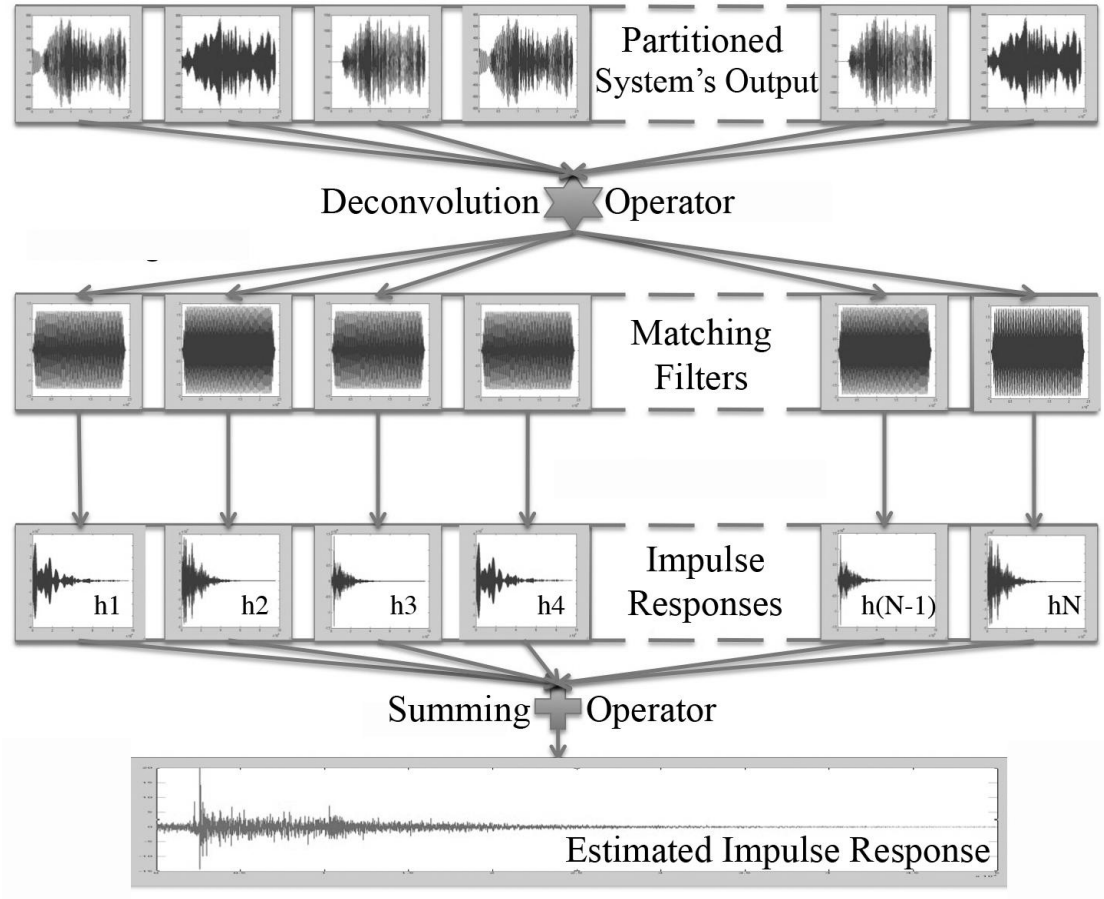


Figure 14 – Framework of the measurements process of the presto-chirps method.

To see how the measurement scheme is practically implemented, a chromatic scale formed by second long presto-chirps spaced apart by two seconds of silence is chosen as an example of a stimulus (this example does not involve musicality), and used to measure a room impulse response $h[n]$(not shown). Figure 15 shows, in the upper panel, the stimulus in a schematic time-frequency representation. The response of the room to the stimulus is instead shown in the bottom panel.

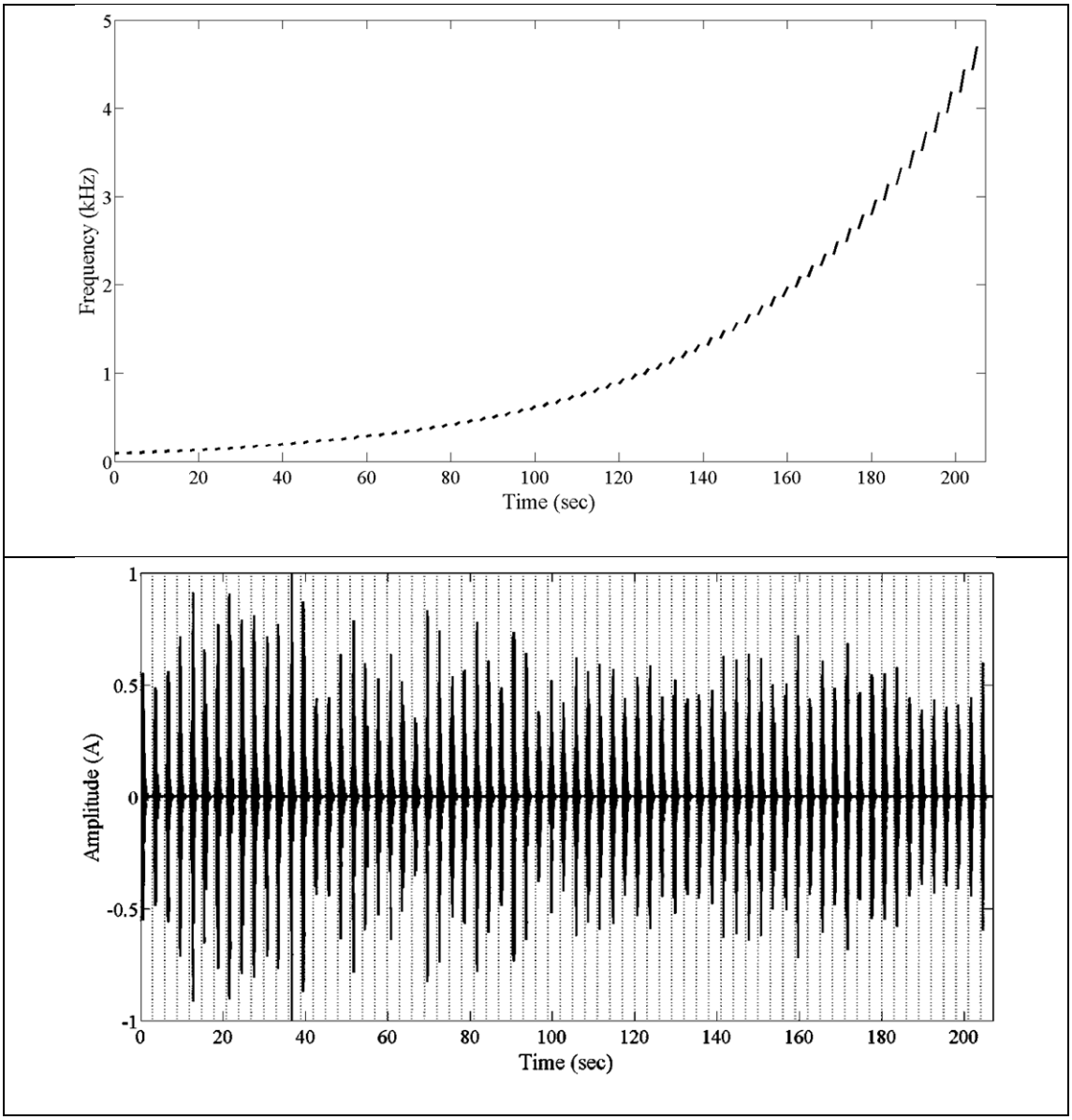


Figure 15 – Upper panel: Stimulus composed of one second-long presto-chirps plus 2 sec of silence. Bottom panel: Recorded response of the room due to the stimulus.

As for the measurement scheme, the output of the system $y[n]$ has to be sliced into chunks, which in this case corresponded to the vertical dashed lines. The length of each chunk is equal to the length of the presto-chirp played in that chunk plus an analysis window (AW), which has to be long enough to contain the whole, or wanted reverberation tail of the room. Each chunk, hereafter denoted by $y_i[n]$, is then deconvolved using the time reversal version of the presto-chirps present in that chunk, that is $\hat{x}_{i,\alpha}[-n]$,

$$h_i[n] = y_i[n] * \hat{x}_{i,\alpha}[-n] \tag{37}$$

where $h_i[n]$ are band limited impulse response. The final step consists on weighting, and sums all the measured partial impulse responses to form the broadband impulse response $h[n]$,

$$h[n] = E_1 h_1[\mathrm{n}] + \cdots + + E_M h_M[n] = \sum_{i=1}^{M} E_\mathrm{i} h_i[\mathrm{n}] \tag{38}$$

where E_i is a weighted factor that is used to compensate for the energy differences of presto-chirps, due to different bandwidths, durations, amplitudes and repetitions, as will be further explained in paragraph 3.5.

3.3 Deconvolution Residuals

The deconvolution of the chunks of system's output $y_i[n]$ can give rise to the "appearance" of artefacts in the measured RIRs, which are inherent to how the deconvolution is performed, and how the presto-chirps are arranged in the stimuli.

For example, the stimulus $s[n]$ is composed of a sequence of N presto-chirps ($\hat{x}_{i,\alpha}[n]$), played at specific time intervals (defined by the sample offset $\mathrm{K}_i \geq 0$),

$$s[n] = \hat{x}_{1,\alpha}[n + \mathrm{K}_1] + \cdots + \hat{x}_{M,\alpha}[n + \mathrm{K}_M] = \sum_{i=1}^{M} \hat{x}_{i,\alpha}[n + \mathrm{K}_i] \tag{39}$$

Feeding $s[n]$ at the input of a system, with impulse response $h[n]$, gives the output

$$y[n] = s[n] * h[n] = h[n] * \sum_{i=1}^{M} \hat{x}_{i,\alpha}[n + \mathrm{K}_i] = \sum_{i=1}^{M} y_i[n + \mathrm{K}_i] \tag{40}$$

Since presto-chirps are used to mimic music, it may happen that $y_i[n]$ contains more than one presto-chirp. This is because notes can be played simultaneously (like in chords) or sequentially. For both situations, it can also happen that the presto-chirps contained in $y_i[n]$ have "common" frequencies. Figure 16 points out such an eventuality as it shows the magnitude responses of two adjacent presto-chirps (for $\alpha = 0.25, 0.5, 0.75, \text{and } 1$). It can be seen that the transition bands of presto-chirps increasingly overlap as α increases.

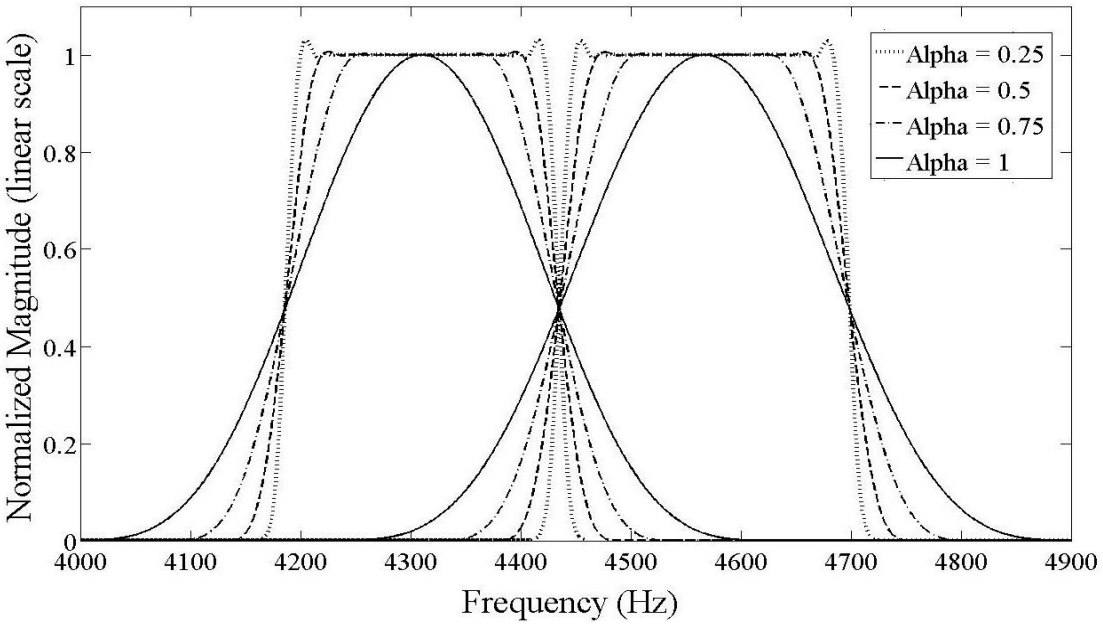


Figure 16 - Normalized magnitude, in linear scale, of two adjacent presto-chirps for different α values of the Tukey windows applied on the presto-chirps.

To supply a practical example of a situation when frequency overlapping could possibly happen, the following is a mathematical analysis of the deconvolution of a chunk of audio containing two presto-chirps. Let

$$y_i[n] = \left(\hat{x}_{i,\alpha}[n] + \hat{x}_{i+1,\alpha}[n]\right) * h[n] \tag{41}$$

where $h[n]$ is the impulse response of the system under test. To obtain the two partial impulse responses $h_i[n]$ and $h_{i+1}[n]$, $y_i[n]$ have to be convolved whether with the inverse filters $\hat{x}_{i,\alpha}[-n]$, and with $\hat{x}_{i+1,\alpha}[-n]$. Thus (in the following the subscript α and the cap are omitted for simplicity of writing),

$$\begin{cases} h_i[n] = y_i[n] * x_i[-n] = \left((x_i[n] + x_{i+1}[n]) * h[n]\right) * x_i[-n] \\ h_{i+1}[n] = y_i[n] * x_{i+1}[-n] = \left((x_i[n] + x_{i+1}[n]) * h[n]\right) * x_{i+1}[-n] \end{cases} \tag{42}$$

which leads to

$$\begin{cases} h_i[n] = \{(x_i[n] * x_i[-n]) + (x_{i+1}[n] * x_i[-n])\} * h[n] \\ h_{i+1}[n] = \{(x_i[n] * x_{i+1}[-n]) + (x_{i+1}[n] * x_{i+1}[-n])\} * h[n] \end{cases} \quad (43)$$

posing $\mathrm{r}_{\mathrm{x}_i}[n] = x_i[n] * x_i[-n]$, and $\mathrm{r}_{\mathrm{x}_{i+1}}[n] = x_{i+1}[n] * x_{i+1}[-n]$, Equation (43) can be rewritten as

$$\begin{cases} h_i[n] = \{\mathrm{r}_{\mathrm{x}_i}[n] + (x_{i+1}[n] * x_i[-n])\} * h[n] \\ h_{i+1}[n] = \{(x_i[n] * x_{i+1}[-n]) + \mathrm{r}_{\mathrm{x}_{i+1}}[n]\} * h[n] \end{cases} \quad (44)$$

The terms $\mathrm{r}_{\mathrm{x}_i}[n]$ and $\mathrm{r}_{\mathrm{x}_{i+1}}[n]$ are the autocorrelation functions of $x_i[n]$ and $x_{i+1}[n]$, respectively. The terms $(x_{i+1}[n] * x_i[-n]) = h_{res1}[n]$ and $(x_i[n] * x_{i+1}[-n]) = h_{res2}[n]$, are residual terms related to the convolution of an inverse filter with an "extraneous" presto-chirp. Rewriting Equation (44) as

$$\begin{cases} h_i[n] = (\mathrm{r}_{\mathrm{x}_i}[n] * h[n]) + h_{res1}[n] \\ h_{i+1}[n] = (\mathrm{r}_{\mathrm{x}_{i+1}}[n] * h[n]) + h_{res2}[n] \end{cases} \quad (45)$$

and adding together the two partial impulse responses $h_i[n]$ and $h_{i+1}[n]$, results in

$$h_{(i,i+1)}[n] = (h_i[n] + h_{i+1}[n]) + h_{res1}[n] + h_{res2}[n] \quad (46)$$

thus, the (summed) impulse responses plus the residual terms. Figure 17 reports a graphic representation of this result. The black intersection area marks the overlapped

transition bands of the two presto-chirps (left panel), and the residuals are shown in the right panel within the summed impulse responses.

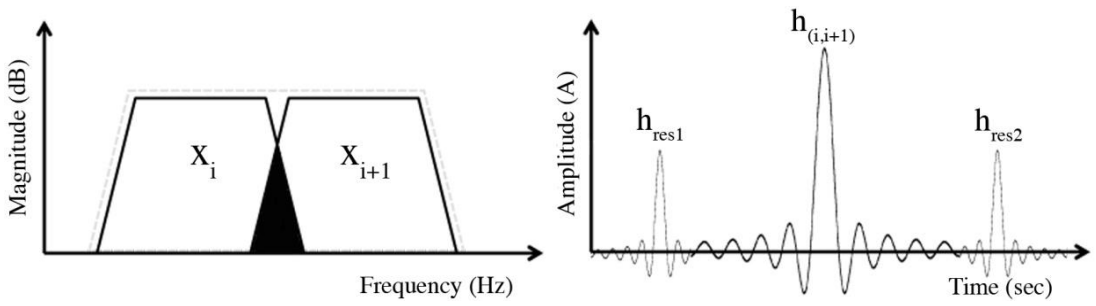


Figure 17 – Superimposed spectra of two adjacent presto-chirps (left panel), and impulse response plus residuals (right panel).

The residual terms, $h_{res1}[n] + h_{res2}[n]$, are zero (or negligible) if there are not common frequencies between the two deconvolved presto-chirps, but will have specific amplitude and time lag otherwise. The shapes and positions of residuals along the RIR depend on both the amount of overlapped frequencies, and on the difference between the instantaneous frequencies of the presto-chirps. In turn, these characteristics depends on the specific Tukey window configuration used, and therefore on α. For example, Figure 18 shows a hypothetical situation, where two frequency adjacent presto-chirps (0.5 sec long, α=1) are deconvolved.

To further extend the discussion, it has to be pointed out that when presto-chirps are windowed using a Hanning window (α = 1), also "two semitones – one tone" apart presto-chirps overlap up to the point of creating residuals. The reason can be seen in Figure 19, which reports the spectra of three adjacent presto-chirps. The transition bands of the two external presto-chirps, for α = 1, overlap around the frequency of 4.3

kHz. Figure 20 shows the deconvolution of a stimulus formed by that triplet of Hanning-windowed presto-chirps. For this case, another pair of residuals is generated.

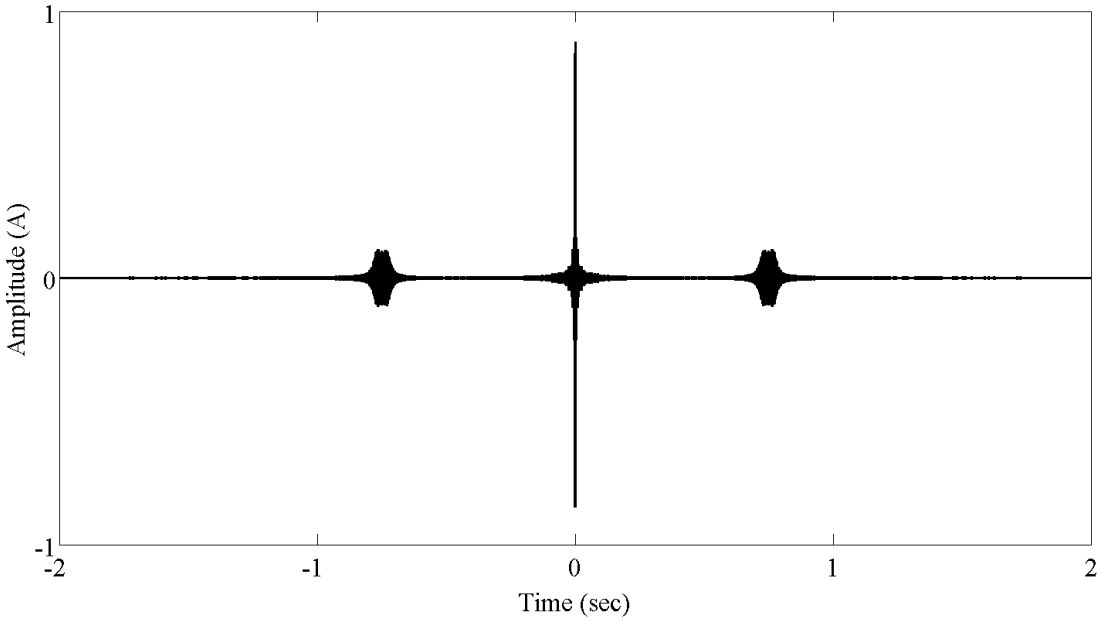


Figure 18 – Example of the appearance of residuals, along the IR, as the result of the deconvolution of two frequency adjacent presto-chirps, in a same chunk of output.

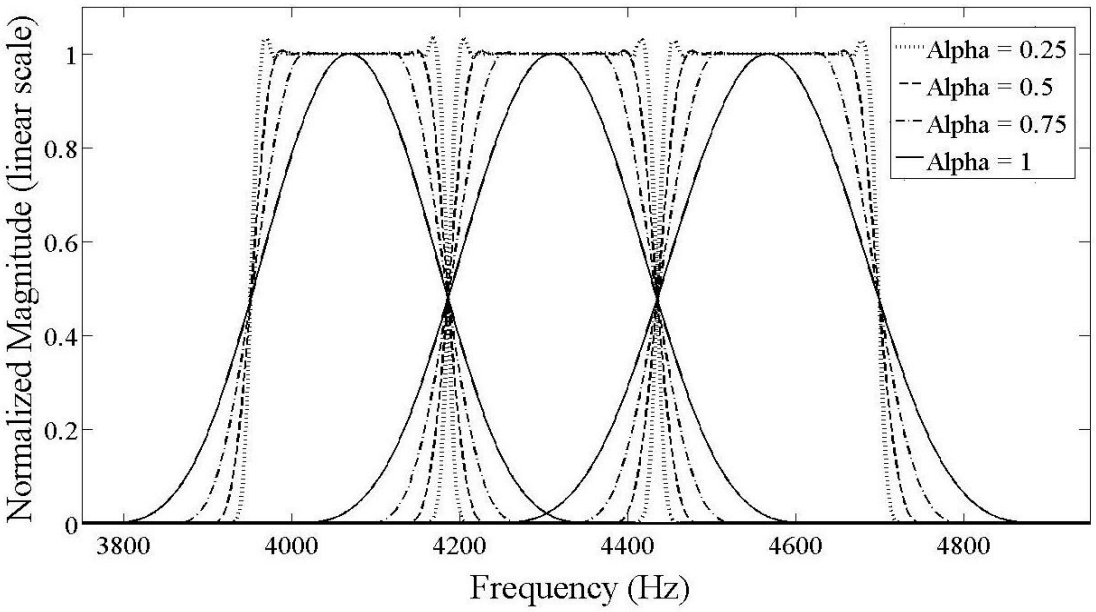


Figure 19 – Normalized magnitude of three frequency adjacent presto-chirps.

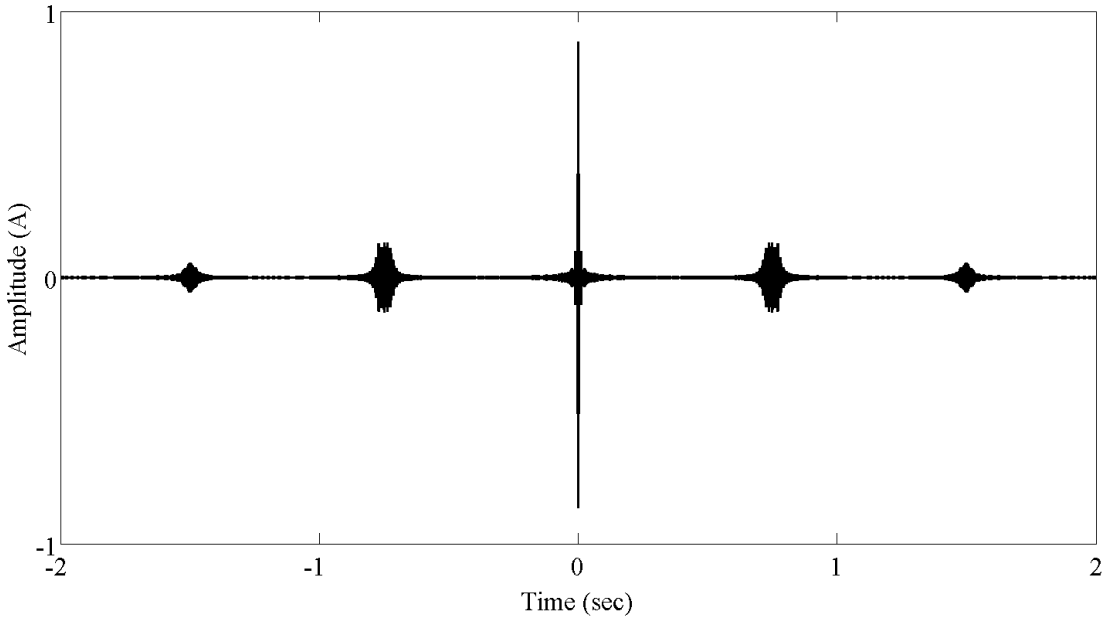


Figure 20 – Residuals as result of the deconvolution of three frequency adjacent presto-chirps for α = 1.

The residuals are deconvolution related artefacts that should be avoided not to compromise the results of the measurements. To avoid the residuals in measured RIR, stimuli should not include repeated or frequency-overlapping presto-chirps played in a time shorter than the length of the RIR to be measured. However, this limitation only exists when presto-chirps have common frequency within a same chunk of audio.

Residuals will be recalled again in the following chapters when other characteristics of presto-chirps, e.g. their duration, will be analysed (Section 4.2.1.5). It will also be discussed how compositions could be made to avoid the formation of residuals, and how this will relate the usability of a stimulus to the length of the RIRs that it can measure (e.g. Section 3.6).

3.4 Measurements Bias

In Section 3.1.2 it has been discussed that Eq. (36) minimises the frequency "gaps" between presto-chirps. Nonetheless, due to a still non-perfect overlapping of the partial impulse responses, a "noise" will bias the measured RIRs. Such bias can be quantified by calculating the frequency response of the sum of the partial impulse responses. For example, using Eq. (37) the deconvolution process can be written as

$$h_i[n] = \left(h[n] * \hat{x}_{i,\alpha}[n]\right) * \hat{x}_{i,\alpha}[-n] = h[n] * r_{i,\alpha}[n] \tag{47}$$

where the term $r_{i,\alpha}[n] = \hat{x}_{i,\alpha}[n] * \hat{x}_{i,\alpha}[-n]$ is the autocorrelation function of the windowed presto-chirps $\hat{x}_{i,\alpha}[n]$. From a signal processing point of view, the deconvolution process can be regarded as a forward-backward filtering of $h[n]$, by the filters $r_{i,\alpha}[n]$, which are zero-phase pass band filters, with order of twice minus one the length of $\hat{x}_{i,\alpha}[n]$. All the M=89 terms $r_{i,\alpha}[n]$ constitute a filter bank (of autocorrelation functions), in which synthesis frequency response is equal to

$$\mathrm{r}_\alpha[n] = \sum_{\mathrm{i}=1}^{\mathrm{M}=89} r_{i,\alpha}[n] \tag{48}$$

Using Equation (48) in Equation (38), results in

$$n] = \sum_{\mathrm{i}=1}^{\mathrm{M}=89} h_i[n] = \sum_{\mathrm{i}=1}^{\mathrm{M}=89} \left(h[n] * r_{i,\alpha}[n]\right) = h[n] * \sum_{\mathrm{i}=1}^{\mathrm{M}=89} r_{i,\alpha}[n] = h[n] * \mathrm{r}_\alpha[n] \tag{49}$$

Ideally $r_\alpha[n]$ should have the impulse response of an ideal band pass filter (from 27.5 Hz to 4.698 Hz), otherwise it will bias the impulse responses. However, the Fourier transform of $r_\alpha[n]$, which is

$$S_\alpha(f) = \mathcal{F}(r_\alpha[n]) \tag{50}$$

Figure 21 shows Eq. (50) for α = 1. The logarithmic magnitude drop is double that in the case of the frequency responses of the set of (Hanning windowed) presto-chirps shown in Figure 13 (because $r_{i,\alpha}[n]$ has a squared magnitude compared with $\hat{x}_{i,\alpha}[n]$). It has to be noted that the magnitude response of $S_{\alpha=1}(f)$ presents a ringing behaviour, which practically is the bias due to a non-perfect recombination of the presto-chirps.

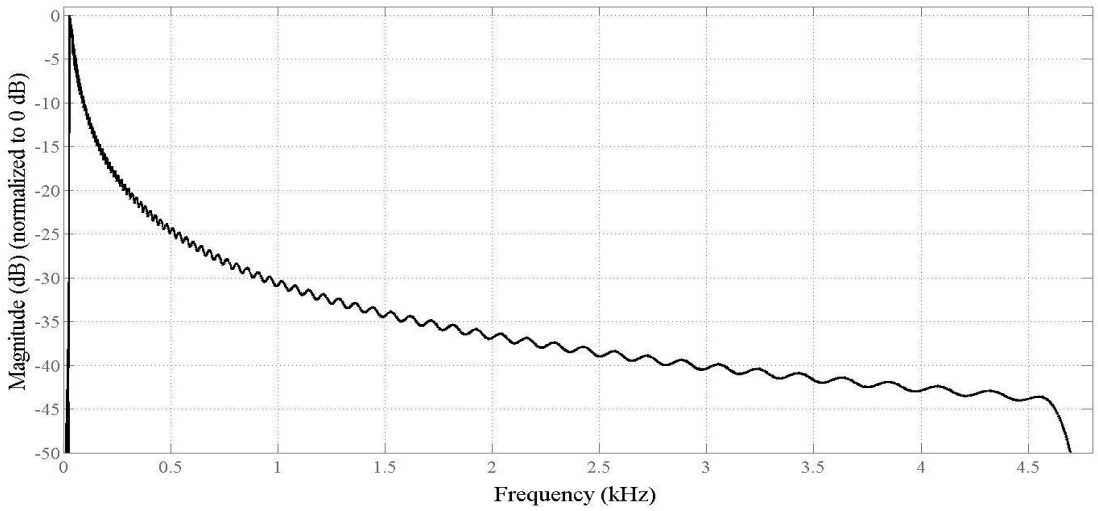


Figure 21 – $S_{\alpha=1}(f)$ of a stimulus formed by the set of 89 presto-chirps.

As mentioned in 3.2 Eq. (38) to achieve a flat frequency response a weighting factor E_i must be used. Part of this factor is made up by the term $BF_{i,\alpha}$, which when multiplied to the partial impulse response $h_i[n]$, or equivalently to the inverse filters $\hat{x}_{i,\alpha}[-n]$ (as the convolution operator possesses the associative property), normalise

the energy differences between presto-chirps due to their different bandwidths. In fact, the term $BF_{i,\alpha}$ is defined as the ratio of the bandwidth of the umpteenth presto-chirp and the minimum bandwidth among the whole set of presto-chirps. Thus

$$BF_{i,\alpha} = B_{i,\alpha}/B_{min} \tag{51}$$

where $B_{i,\alpha} = f_{i+1,\alpha} - f_{i,\alpha}$ is the bandwidth of a presto-chirp calculated using the "new" frequencies as for Equation (36), and $B_{min} = \min_{\mathrm{i}}\{\mathrm{B}_{i,\alpha}\}$.

Figure 22, Figure 23, Figure 24, and Figure 25, shows the results of such normalisation for $\alpha = 1, 0.75, 0.5, \text{and } 0.25$, respectively; in the upper panels are overlaid the spectra of the terms $r_{i,\alpha}[n] \cdot BF_{i,\alpha} = \hat{x}_{i,\alpha}[n] * \left(BF_{i,\alpha} \cdot \hat{x}_{i,\alpha}[-n]\right)$, whilst in the bottom panels are reported the (normalised) synthesised frequency responses

$$\bar{\mathrm{S}}_{\alpha}(f) = \mathcal{F}\left\{\sum_{\mathrm{i}=1}^{\mathrm{M}}\left(\, r_{i,\alpha}[n] \cdot BF_{i,\alpha}\right)\right\} \tag{52}$$

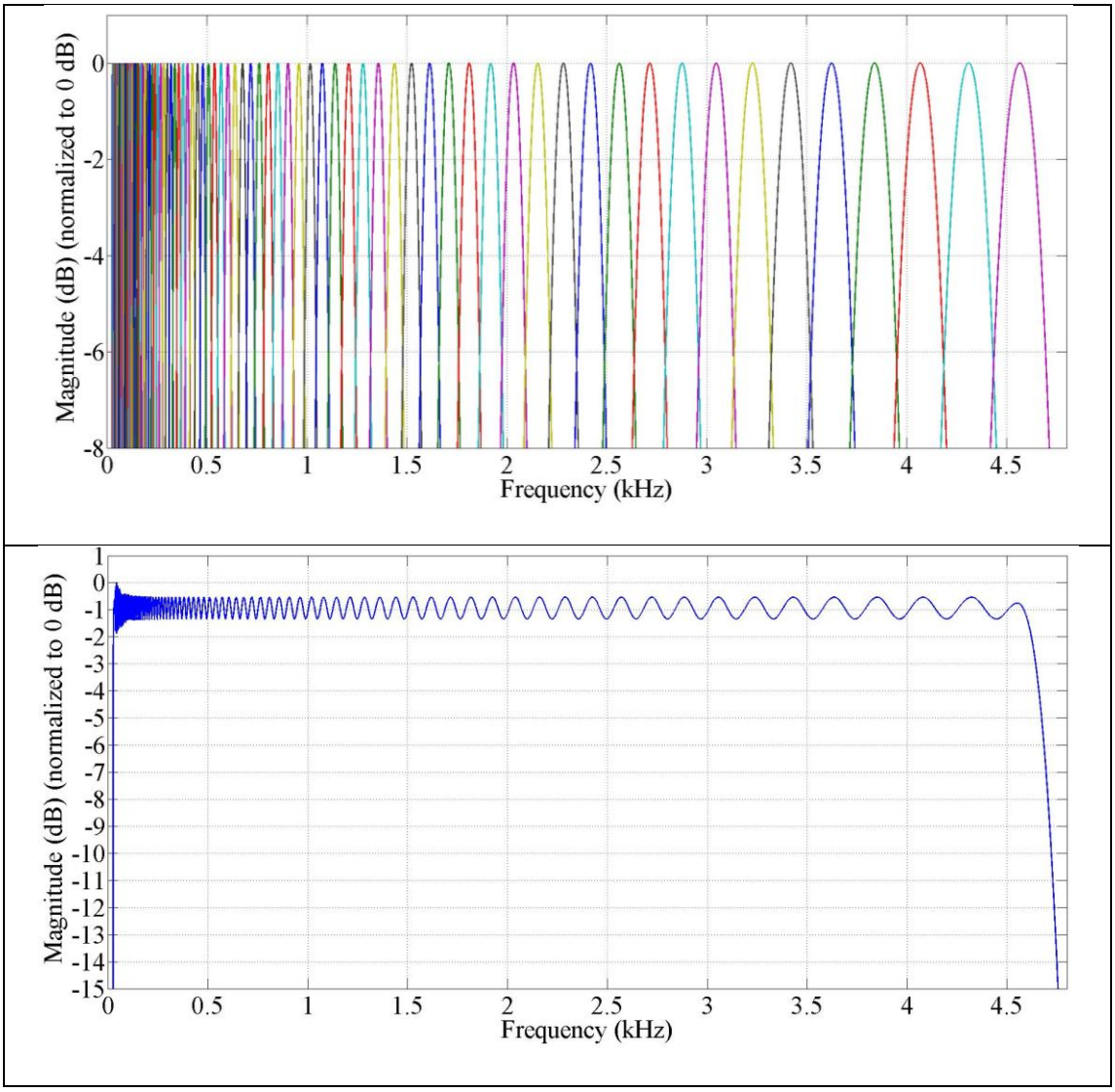


Figure 22 – Upper panel: overlaid magnitude responses of the 89 $r_{i,\alpha=1}[n]$. Bottom panel: normalised synthesised frequency response of the set of presto-chirps $\bar{S}_{\alpha=1}(f)$.

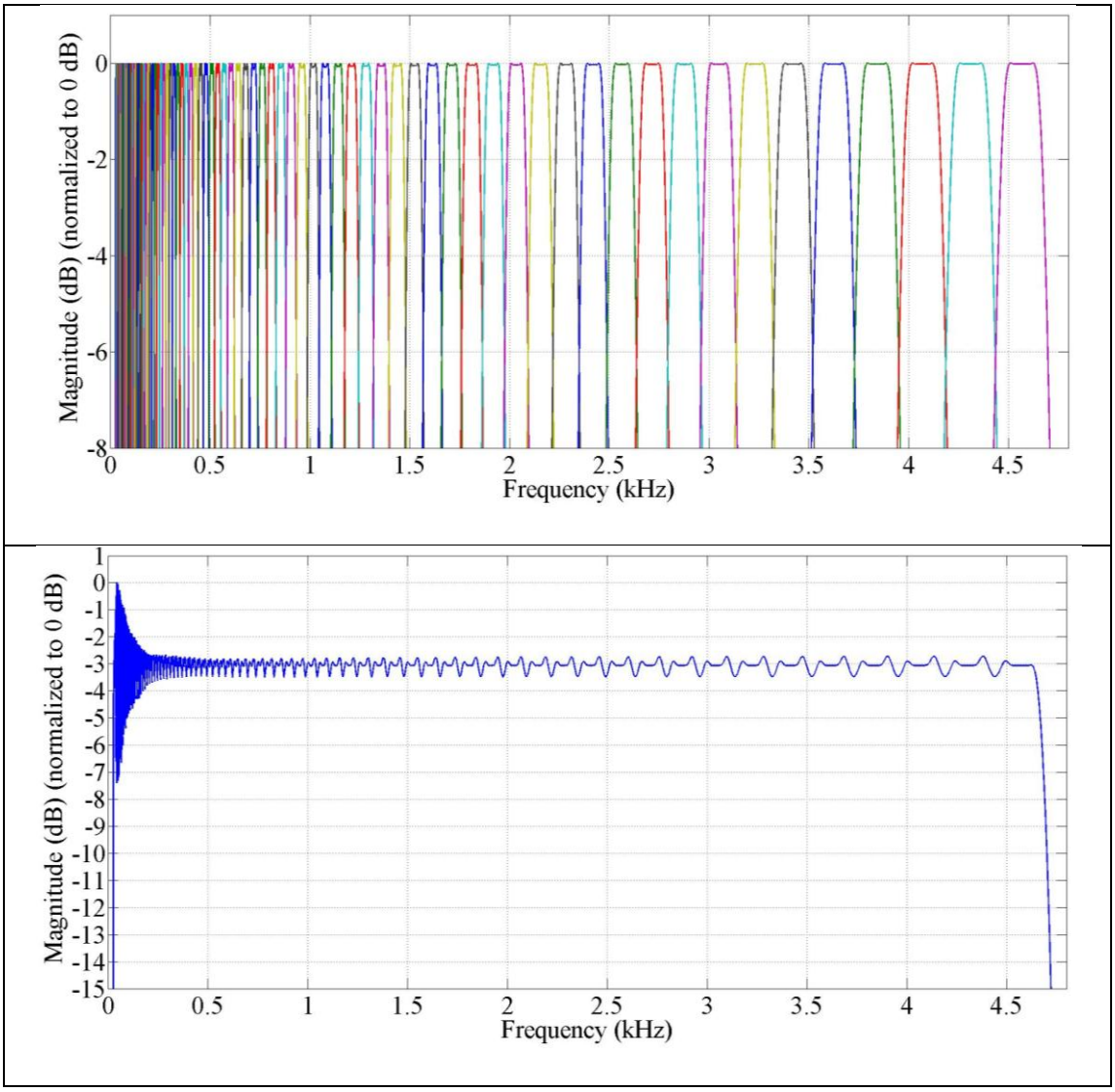


Figure 23 – Upper panel: overlaid magnitude responses of the 89 $r_{i,\alpha=0.75}[n]$. Bottom panel: normalised synthesised frequency response of the set of presto-chirps $\bar{S}_{\alpha=0.75}(f)$.

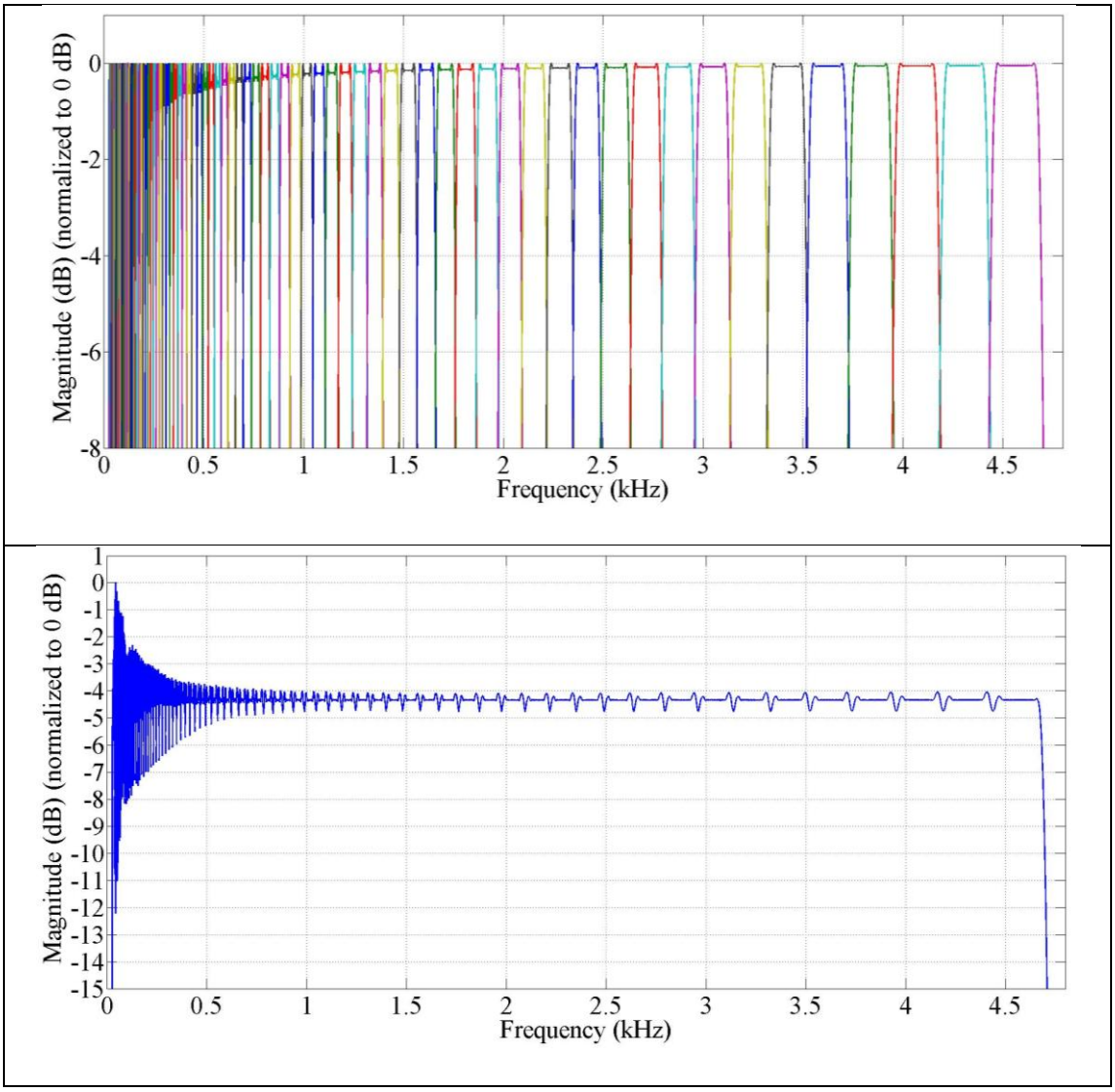


Figure 24 – Upper panel: overlaid magnitude responses of the 89 $r_{i,\alpha=0.5}[n]$. Bottom panel: normalised synthesised frequency response of the set of presto-chirps $\bar{S}_{\alpha=0.5}(f)$.

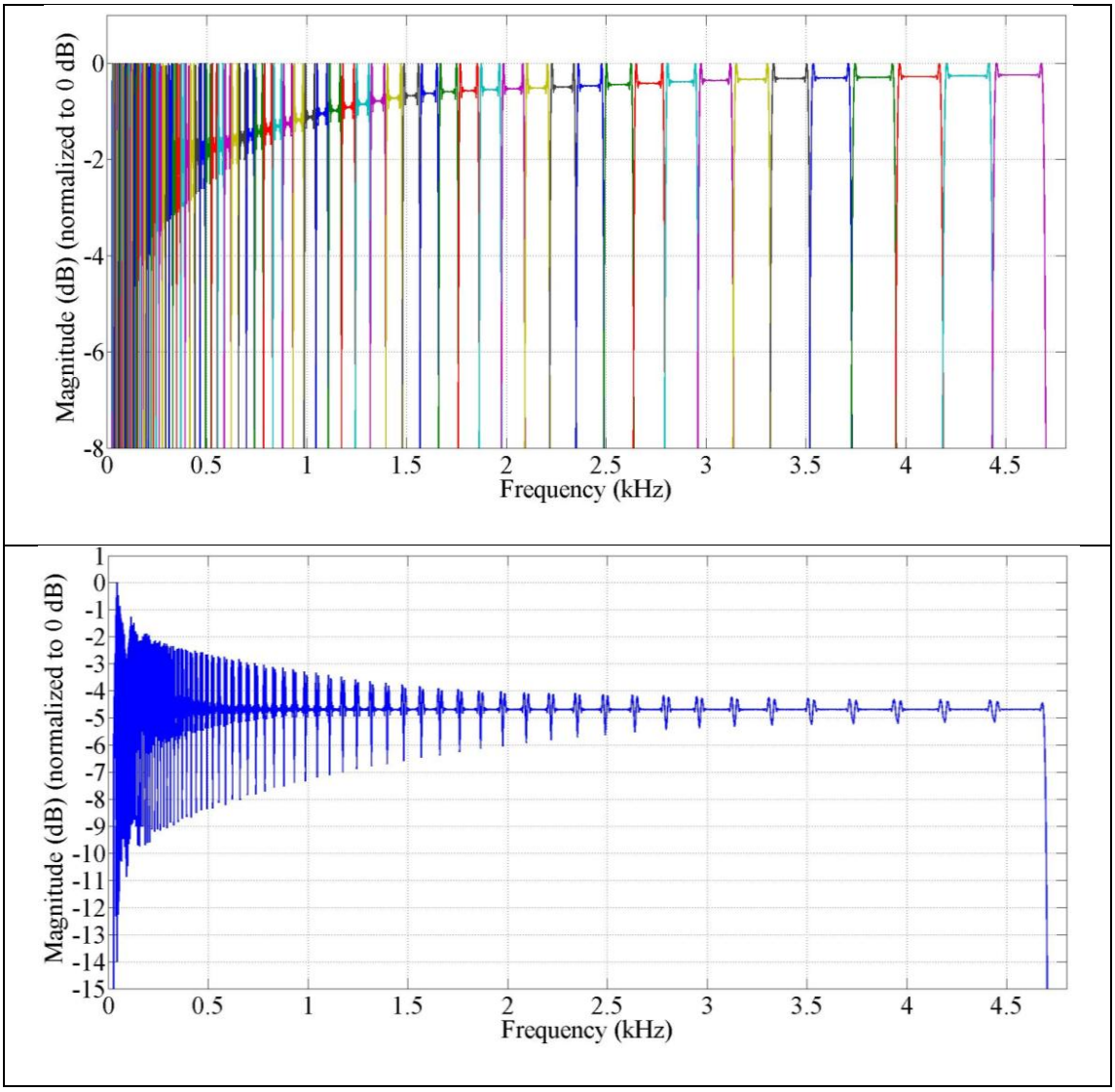


Figure 25 – Upper panel: overlaid magnitude responses of the 89 $r_{i,\alpha=0.25}[n]$. Bottom panel: normalised synthesised frequency response of the set of presto-chirps $\bar{S}_{\alpha=0.25}(f)$.

In the bottom panels of the previous figures it can be seen that, depending on which Tukey window configuration is used, the cumulative frequency response shows more or less unevenness. The worst-case scenario is for $\alpha = 0.25$ at low frequencies. The reason is the minor attenuation of the ripples at the frequency edge of presto-chirps, which is more pronounced at low frequencies (as can be seen in the upper panels of the same figures), which in turn is related to the minor bandwidth of presto-chirps at low frequencies.

In practice, the best result, in term of ripples attenuation, and low magnitude "irregularities" in the cumulative frequency response, is obtained by using a Hanning window (case α = 1), as can be seen in Figure 22. For this configuration, the bias introduced in the frequency response is of about 1 dB for the whole spectrum, until almost 100 Hz where the bias increases up to 2 dB. However, in Section 4.2.1.5 it will be shown than the magnitude of the bias also depends on the duration of the presto-chirps.

In general, it can be concluded that the use of Hanning windowed presto-chirps will introduce the minimal amount of bias, in the measured RIR, compared with the use of the other tested window configurations. This conclusion will be further validated in chapter 4.

3.5 Normalisation and Final Sum

Within a score, presto-chirps are likely to have different durations and amplitudes as they follow a rhythmic and melodic pattern. This leads to presto-chirps being fed with different energies into the room under test. Consequently, the measured partial impulse responses will have different energy (different SNR too). To normalise such differences, which otherwise will comport a frequency coloration of the measured RIR, each partial impulse response $h_i[n]$ is divided by the energy of the presto-chirps $\hat{x}_{i,\alpha}[n]$, thus their autocorrelation function at zero lags $ACF(\hat{x}_{i,\alpha}[n])$. This normalisation, which is common in chirp-based methods, also compensates for the introduced magnitude-squared energy of estimated impulse.

In addition, since presto-chirps could be used more than once within a score, the "duplicate" copies of $h_i[n]$ that they will produce have to be averaged together before the final sum (again to avoid frequency coloration of the measured RIRs). This is achieved by dividing each $h_i[n]$ by a number indicating the times it has been measured (rep_i). It is worth noting that averaging impulse responses leads to an increase of the SNR in specific sub bands and that this can be used to address noisy sub bands. Moreover, it permits the discarding of those duplicated $h_i[n]$, which have been "tainted" by noisy events, and/or residuals. Summarising all the weighting factors discussed so far, the term E_{i} presented in Eq. (38), results equal to

$$E_{\mathrm{i}} = \left(\frac{BF_{i,\alpha}}{ACF(\hat{x}_{i,\alpha}[n])}\right) \cdot \frac{1}{rep_i} \tag{53}$$

where, the parameter BF_i is used to normalise the different presto-chirps' bandwidths, the parameter $ACF[x_i[0]]$ accounts for presto-chirps' different lengths and amplitude (and the squared energy), and the parameter rep_i computes the average of repeated impulse responses $h_i[n]$.

Finally, expanding Eq. (38) using Eq. (53) results in the final formula used to sum and weight the measured partial impulse responses, which then leads to the sought broadband RIR. That is

$$\begin{aligned} h[n] &= \left(\frac{BF_{1,\alpha}}{ACF(\hat{x}_{1,\alpha}[n])}\right)\cdot\frac{h_1[n]}{rep_1} + \cdots + \left(\frac{BF_{M,\alpha}}{ACF(\hat{x}_{M,\alpha}[n])}\right)\cdot\frac{h_M[n]}{rep_M} \\ &= \sum_{i=1}^{M}\left(\frac{BF_{i,\alpha}}{ACF(\hat{x}_{i,\alpha}[n])}\right)\cdot\frac{h_i[n]}{rep_i} \end{aligned} \tag{54}$$

Where M is the number of the presto-chirps used in the measurements, which may coincide or not with the actual notes used in stimuli, as they may be used only to improve the stimuli musicality (and they may produce residuals and therefore need to be discarded). It can also be debated that other "sounds" can be mixed with presto-chirps to embellish the compositions, providing they do not cause deconvolution residuals. However, it is a fact that M cannot be less than 89 since all presto-chirps have to be used at least once to cover the whole frequency range.

3.6 Presto-chirps Stimuli

Presto-chirps are constructed around the frequency of musical notes (equal tempered chromatic scale) to prove experimentally if they can be used to compose unobtrusive stimuli, to be used for occupied measurements. The idea that has been thought of as straightforward to create unobtrusive stimuli was to use existing music compositions. For the sake of argument, it must be clarified that this thesis aims at demonstrating that presto-chirps could "eventually" be transformed into well-sounding music-like stimuli. Therefore, the effort that has been put in this research in composing musical stimuli is chiefly oriented in delimiting their usability, and reliability, in performing acoustic measurements. This consequentially highlights the limitations of the presented work on resolving the artistic trait of stimuli composed using presto-chirps.

Nonetheless, in the following, the relation between presto-chirps stimuli and the measurable length of RIRs is discussed first. Secondly, is detailed a developed procedure to help create, modify, and subsequently use the presto-chirps. Thirdly, is presented the presto-chirps composition, readapted from an existing song, which has been used to validate the in situ measurements.

3.6.1 Measureable RIR Length

Residuals, as it has been discussed in 3.3, can be produced by the presence of presto-chirps with common frequencies if they are within the same block of deconvolved audio. The time-frequency distance between presto-chirps (thus the difference between their instantaneous frequencies) and the amount and magnitude of the overlap (which depends on the specific window used) determine where residuals will

form in the measured RIRs. Calculating analytically the positions of residuals is difficult, However, simulating the measurement of a known RIR using a composed stimulus supplies an easier solution to the problem. Based on the results from the simulation, the found distance from the main peak of the measured RIR and the first residuals, defines the length of the analysis window that can be use to deconvolve the acoustic response of a system for that stimulus.

Practically, the analysis window defines where the measured RIR is truncated, thus the overall measured length. To measure longer RIRs, it could be possible to "re-arrange" a stimulus, or creating others anew. Re-arranging a stimulus implies first searching for those situations that will produce residuals. This, from the point of view of music composition, corresponds to notes with a critical "semitone distance" to one another, played during an interval of time as long as the used analysis window.

Such semitone distance is related to the length of the presto-chirps, and of the specific Tukey-window applied to the presto-chirps. Generally, it is about two semitones for $\alpha \neq 1$, and 3 semitones for $\alpha = 1$, and doubles for presto-chirps shorter than about 1 second for frequencies below about 100 Hz (reasons are given in Section 4.2.1.5). Moreover, software simulation, as further discussed in Appendix A, are conveniently used to determine accurately the presence of residuals for given stimuli.

Avoiding the use of notes within some semitone distances, although limits the free composing of music (which however has its own rules), it guarantees the non-appearance of residuals in the (truncated) RIR. Moreover, it can be argued that since a song can be composed by an arbitrary number of notes, it suffices to discard, from the

final sum (Eq. (54)), those partial RIRs that present residuals to address the problem. This, however, is if the frequency range of interest is fully covered. Consequently, the "forbade" cases can still be used to accommodate the musicality of stimuli.

3.6.2 Structured Presto-chirps Stimuli

To facilitate using presto-chirps, the MIDI notation is conveniently used to represent them, and MIDI files used to store composed stimuli. A MIDI file is an "organised table" whose rows represent the notes as listed in a music score, and whose columns contain information about each note, including the information used to control MIDI compatible instruments. MIDI stands for Musical Instrument Digital Interface, and its introduction and use on the market followed the developing of electronic instruments. The advent of software for music composition boosted MIDI files to become largely used for storing music, which was also prompted by the possibility that they offer the ability to be played by "virtual" instruments. Noting that, for example, presto-chirps could be played using a virtual keyboard where presto-chirps are associated to its keys.

Developed software routines have been used to access MIDI files and retrieve each note of a song, its amplitude, pitch and the playing times (thus when it starts and ends in the song). Other purposely-developed functions use those parameters to create presto-chirps stimuli, which are then converted into an audio file, which is actually used for the measurements. The deconvolution process is then performed by retrieving the information stored in the files, which are used to create the inverse filters needed to deconvolve the "blocks" of the system output. The obtained partial impulse responses are summed up, as described in Eq. (54), to form the sought RIR.

MIDI files conveniently allow the use of music notation software to create and/or modify music compositions. In fact, theoretically, any existing song could be transformed into presto-chirps stimuli. However, as it has resulted, most songs do not actually use all the semitones the presto-chirps methods needs, and although a "medley" of several songs could be used to cover more semitones, this approach has proved cumbersome and ineffective. In addition, since presto-chirps are "glissando" notes, and only made up of fundamental frequencies, even if they are used to mimic existing music the resulting melodies will be quite different compared from the original versions.

3.6.3 Used (Music-like) Stimulus

After a long search, few musical compositions have been found that use all the 88 notes of a grand piano. The contacted composers have kindly allowed the use of their music for the scope of this research, also supplying the MIDI files. From those few, one song has been chosen based on personal judgement of how it sounded once transformed into a presto-chirps stimulus, and it has been used to carry out in situ measurements. The song is named after a famous road in Rome, "Via del Corso", and written by the Finish composer Tero-Pekka Henell. It has been modified, although this compromised its musicality, to ensure that the stimuli could be used without producing convolution residuals up to a length of one second. The duration and amplitude of the highest pitched notes were decreased to avoid the nuisance of high-pitched sounds. The amplitude and duration of the low-pitched notes has been slightly increased to feed more energy at the low sub bands. The 89^{th} note, the highest in pitch, has been added to meet the specificity of the presto-chirp method. These

modifications are reported, for the sake of completeness, in Appendix A. In the following instead Figure 26 reports the stimulus time-(logarithmic)frequency representation, and Figure 27 shows the waveform of the stimulus for Hanning-windowed presto-chirps.

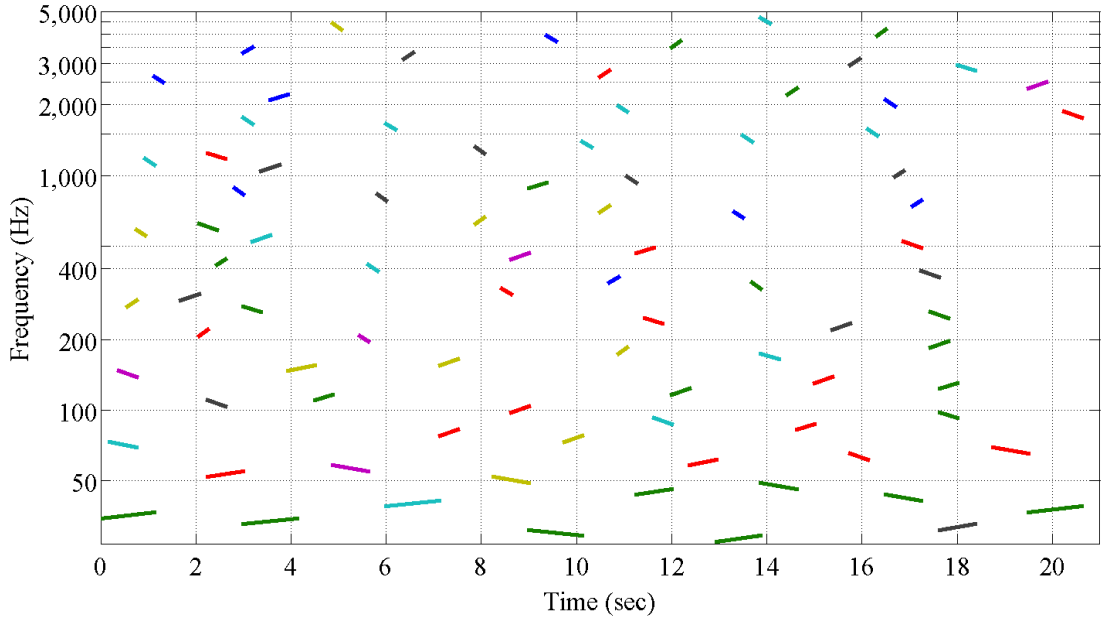


Figure 26 – Time-(log)frequency representation of the stimulus.

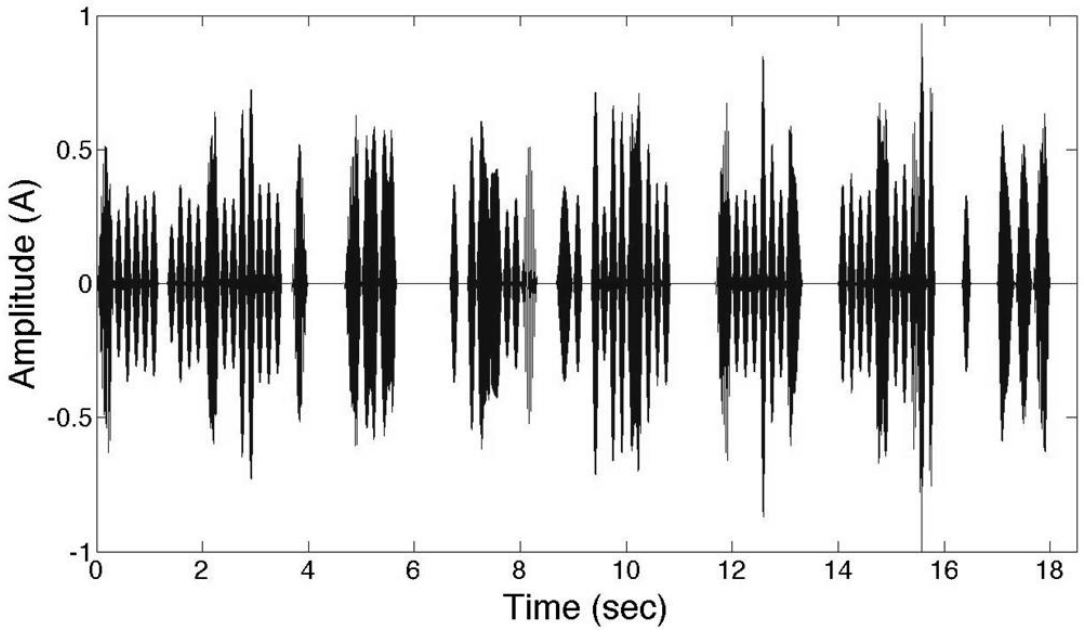


Figure 27 – Waveform of the re-arranged song "via del Corso". Hanning windowed presto-chirps have been used in this representation.

4 Validations of the Presto-chirps Method

Validation measurements have been carried out, in simulated scenarios to verify the method accuracy in time invariant systems (Section 4.2), and in-situ to validate the method in real environments (Section 4.3).

The software simulations were used to perform measurements of "known" impulse responses using presto-chirps stimuli. This extends the theoretical argumentations discussed in Chapter 3. Two scenarios will be presented. The analysis of

- a (FIR) low pass filter.
- a recorded impulse response of a church.

The stimuli used in these tests were, in the first case, a random generated stimulus (formed by 0.125 seconds long presto-chirps), and in the second case, a chromatic scale formed by half second long presto-chirps. The first simulation is used to shed a light on the achievable accuracy for each of the Tukey windows discussed in 3.1.1; both in time and in frequency domain. The use of short chirps (1/8 of a second) is to demonstrate the relation between the duration and the resulting spectra of presto-chirps, and consequently their cumulative frequency response. The second simulation is used to demonstrate the accuracy of the method in dealing with a time invariant room impulse response, since this allows for excluding the presence of external sources of errors in the measurements.

On the other hand, in situ measurements have been carried out to demonstrate the accuracy of the method in dealing with real (time variant) spaces. The measurements were performed using the (musical) stimulus presented in Section 3.6.3, in the following venues:

- Two auditoria (used for live events for radio broadcasting) - where the actual number of participating/attending people can vary from day to day.
- A broadcasting studio room – which can host musicians and a small audience and which numbers vary depending on the specific show.

In situ validations are reported in 4.3. However, since results are very similar, only the results for one of the tested auditorium ("Sala B") are presented in this chapter. The other two cases are reported, for the sake of completeness, in Appendix C.

4.1 Validation Criteria

This section presents the validation criteria used to validate the accuracy of the measurements. They consist of comparisons between measured and referenced acoustic parameters, EDCs, impulse responses (correlation analysis), and frequency responses (coherence analysis). Results are discussed in Sections 4.1.1, 4.1.2, and 4.1.3, respectively. For the two simulated scenarios, point-to-point magnitude and amplitude difference functions are also used to highlight theoretical issues. These functions have not been used in real scenarios since real environments are subjected to fluctuations of their acoustical space, which hinder an objective analysis of the results.

4.1.1 Acoustic Parameters Analysis

The monaural acoustic parameters (described in 2.2) clarity (C_{80}), definition (D_{50}), early decay time (EDT), and reverberation times (RT30, and RT20), are calculated from RIRs measured using the presto-chirps method. They are then compared with "reference" acoustic parameters estimated from the same RIRs, during the same measurement sessions, or in the case of simulated scenario with a known RIR. The acoustic parameters are calculated from the measured RIRs using the plug-in "Aurora" for "Audacity" (Campanini & Farina, 2009), for the octave bands from 31.5 Hz up to 2 kHz, and in the third-octave band centred at 4 kHz. The reference and measured parameters are compared by taking the absolute value of the difference between a reference and measured value. For example, in the case of the parameter RT30, the following formula has been used

$$Diff(RT_{30}) = |RT_{30}(reference) - RT_{30}(measured)| \text{ [sec]} \quad (55)$$

where $RT_{30}(reference)$, and $RT_{30}(measured)$, are the reference and measured RT respectively. The so obtained difference value $Diff(RT_{30})$ is compared against the difference limens (DL) of the parameter. The DL represents the smallest change, or equivalently the just noticeable differences (JND) or the just-not-noticeable difference (JNND) – depending on the testing method used - that can be subjectively detected. Therefore, if the difference between the reference and measured acoustic parameters is smaller than that its DL, it can then be assumed that differences in the measured RIR should not be noticed. The DLs used in this thesis depends if the RT is minor or greater than 0.6 seconds. Table 4 reports the used values and the authors of the research from where those values have been taken.

Acoustic parameter	DLs (**RT < 0.6**	DLs (**RT > 0.6 sec**)
Clarity ($\boldsymbol{C_{80}}$) [dB]	*Not relevant to perception at low frequency*	***1 dB*** (Bork, 2000)
Definition ($\boldsymbol{D_{50}}$) [%]	*Not relevant to perception at low frequency*	***5%*** (Bork, 2000)
Early Decay Time **EDT** [s]	***0.024 sec*** (Seraphim, 1958)	***3.2% ± 0.6%*** (Tsolias & Davies, 2014)
Rev. Time $\boldsymbol{RT_{30}}$ $[s]$	***0.024 sec*** (Seraphim, 1958)	***3.2% ± 0.6%*** (Tsolias & Davies, 2014)
Rev. Time $\boldsymbol{RT_{60}}$ $[s]$	***0.024 sec*** (Seraphim, 1958)	***3.2% ± 0.6%*** (Tsolias & Davies, 2014)

Table 4 - Difference limens used to validate the measurements using presto-chirps stimuli. The rightmost column reports the authors of the research from where the DLs have been extracted.

Note that the acoustic parameters C_{80}, and D_{50} are listed as not relevant to perception in small room at low frequencies. This statement derives from a personal conversation with the one of the author of the used research (W. J. Davies). Although the DLs are generally calculated on broadband measures (thus the mean value of all octave bands), in this thesis, they are calculated in octave bands (third-octave band for the one relative at the 4 kHz sub band), as given by the aforementioned software plug-in "Aurora". On the other hand, the "broadband" reverberation time used to define which DLs had to be used in the comparisons is calculated as the mean value of the third-octave bands from 63 Hz up to 4 kHz.

The analyses of the acoustic parameters are presented for each case study grouped in tables, which are divided into blocks. Each block relates to a specific acoustic parameter, whose name is labelled in the vertical green field along with the used difference limens. The leftmost column of each block indicates which Tukey window configuration has been used in the measurement. The first row of each block reports the reference value of the acoustic parameters (written in blue). For the cases in which the RT was > 0.6 sec, a second row (in green and in Italic) is added to indicate the exact value of the difference limen in relation to the reference value (since DLs are in percentage). The following rows, from top to down, show the calculated difference values for each Tukey windowed configuration, as labelled. Each column, after the first indicates the specific octave bands the shown value refers. The 4 kHz sub band is marked with a star to indicate that it has been calculated in third-octave. The values written in red are greater than their specific difference limen. Table 5 reports an example of such representation, related to the block RT60.

RT60 [sec] (*DL = 3.2% ± 0.6%*)	Bands [Hz]	31.5	63	125	250	500	1k	2k	4k*
	Reference	1.28	1.14	0.82	0.67	0.57	0.41	0.44	0.45
	DL 3% (sec)	*0.04*	*0.03*	*0.02*	*0.02*	*0.02*	*0.01*	*0.01*	*0.01*
	α = 1	0.13	0.03	0.02	0.02	0.00	0.00	0.00	0.00
	α = 0.75	0.22	0.10	0.02	0.01	0.02	0.00	0.01	0.01
	α = 0.5	0.50	0.18	0.07	0.03	0.02	0.01	0.01	0.01
	α = 0.25	0.73	0.28	0.16	0.04	0.12	0.15	0.13	0.11
	α = 0	0.84	0.38	0.20	0.11	0.37	0.27	0.27	0.28

Table 5 – Example of a block relative to the measure of the acoustic parameter RT60.

4.1.2 Time Domain Analysis

To validate the accuracy of the method in the time domain, visual comparisons of the reference and measured RIRs are used, along with a correlation analysis. This latter is used to supply more objective clues about the similarity between the two confronted RIRs. It correlates 10ms long slices of the RIRs, which is about the temporal resolution of human hearing at 4.8 kHz (M. J. Shailer and B. C. J. Moore, 1983). For lower frequencies, the temporal resolution is longer. Figure 28 presents an example of this comparison. Besides the RIRs, a yellow box is used to present the RIRs' early reflections, up to about 50 ms; the green vertical dashed lines indicate the correlation values, which are bounded between one and zero, where one indicates complete correlation.

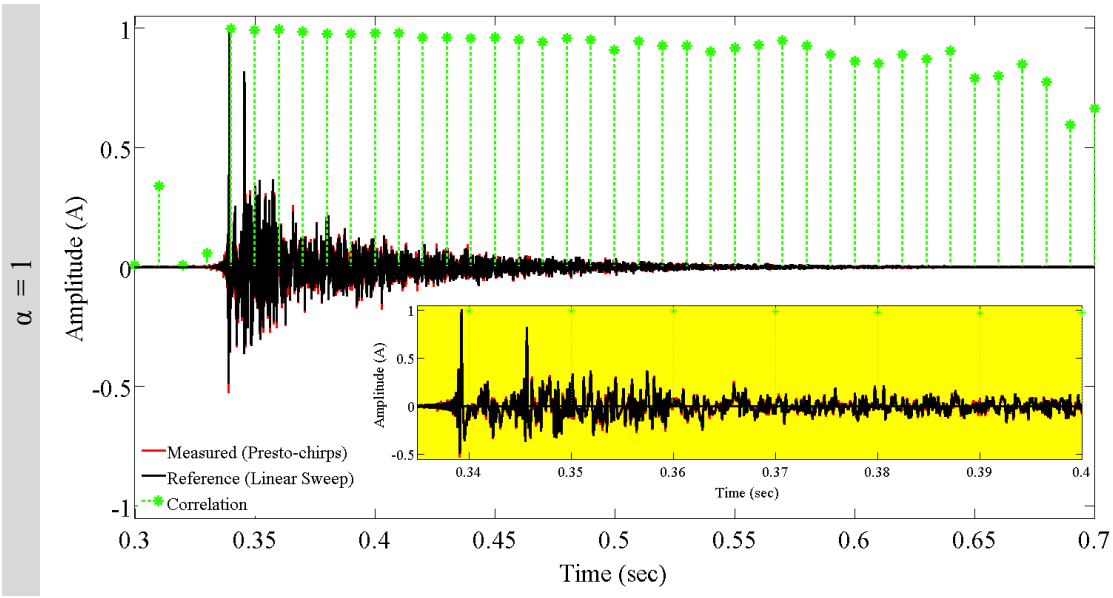


Figure 28 - Example of the visual comparison of RIRs, and its early reflection (yellow box). The green vertical dashed lines mark the correlation between the RIR. Each line represents the correlation value of a 10ms long section of the RIR.

For the simulated scenarios, the absolute value (in dB) of the amplitude difference between the reference and the measured RIR is also presented. Such difference is calculated as

$$Amp_{diff}[n] = 20\log_{10}\left(1 + \left|RIR_{ref}[n] - RIR_{mes}[n]\right|\right) \quad \text{[dB]} \qquad (56)$$

where $Amp_{diff}[n]$ is the amplitude difference, $RIR_{ref}[n]$, and $RIR_{mes}[n]$ are the reference and the measured RIR, respectively.

4.1.3 Frequency Domain Analysis

Frequency domain validations are carried out using the (magnitude squared) coherence, between the impulse response measured using a presto-chirps stimulus and the know RIR. The coherence is calculated using the function “mscohere” in

MATLAB (which is based on the work of (Welch, 1967). The coherence function is a real valued function (thus phase information is discarded), and it is mathematically defined as

$$C_{xy}(f) = \frac{\left|S_{xy}(f)\right|^2}{S_{xx}(f)S_{yy}(f)} \tag{57}$$

where $S_{xx}(f)$, $S_{yy}(f)$, are the auto spectra of the input signal $x[n]$ and the output signal $y[n]$, respectively. $S_{xy}(f)$ is the cross spectrum of the two signals. The coherence represents how much of the spectrum of $y[n]$ is due to the excitation signal $x[n]$. It is bounded in the range 0, and 1, thus $0 \leq C_{xy}(f) \leq 1$, where one stands for total coherence. Lower values indicate the presence of noise or distortions in the output signal. In this validation, the input $x[n]$ is substituted by the reference impulse response $RIR_{ref}[n]$, and the output $y[n]$ by the measured one $RIR_{mes}[n]$. The aim of the coherence analysis is to check the differences between the frequency responses of the two RIRs (given by their autospectra). Coherence measures are presented in figures, at which bottom is also shown the SNR calculated from the coherence using the relation

$$SNR(f) = \frac{C_{xy}(f)}{1 - C_{xy}(f)} \tag{58}$$

Figure 29 presents an example of a coherence analysis, for a measurement taken using Hanning windowed presto-chirps (α=1 as labelled in the left). The vertical green lines indicate the frequencies of musical notes. The coherence is computed using only the

part of the RIR, which is not biased by the background noise, as it will be indicated in the respective analyses, thus, generally, between 0.3 and 0.5 seconds.

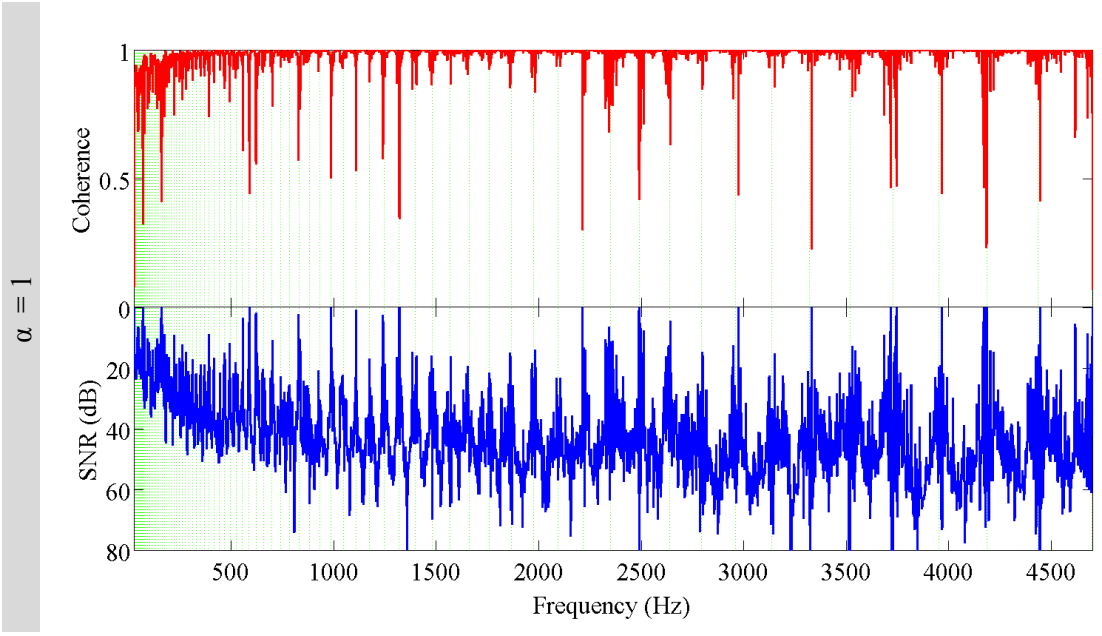


Figure 29 – Example of measured coherence and SNR function, for a measurement carried out using a presto-chirp stimulus windowed with a Tukey with α = 1.

For the two simulated scenarios, the absolute magnitude difference between the spectra of the reference and measured RIRs is also presented. It is defined as

$$MagDiff(f) = 20\log_{10}\left(\frac{\left|\mathcal{F}(RIR_{ref}[n])\right|}{\left|\mathcal{F}(RIR_{mes}[n])\right|}\right) \quad [dB] \tag{59}$$

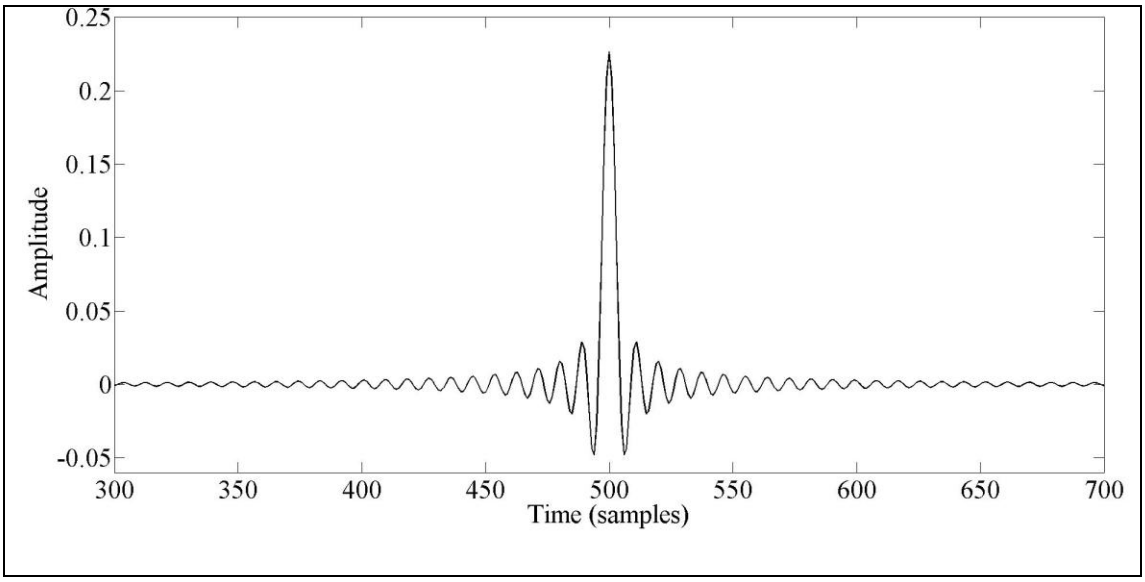
where $\mathcal{F}(RIR_{ref}[n])$ and $\mathcal{F}(RIR_{mes}[n])$ are the Fourier transform of the reference and the measured RIR, respectively. These difference functions are shown, for the frequency range from 27 Hz to 4.7 kHz, in tables whose rows related to the specific Tukey window applied to the presto-chirps as a function of α.

4.2 Computer Simulations

The computer simulations have been implemented in MATLAB. Algorithms have been developed to: read the MIDI file of the stimulus, create its "audio" version, convolve it with the impulse response of the simulated scenario, deconvolve the system output, and then perform the weighted sum of the partial impulse responses.

4.2.1 Case Study: Low Pass Filter

In this test, the impulse response of a 1000 taps low pass filter (fc = 5 kHz) is measured. Figure 30 shows, in the upper panel, its (casual) impulse and, n the bottom panel, its frequency responses (the red line represents the ideal frequency response of such filter).



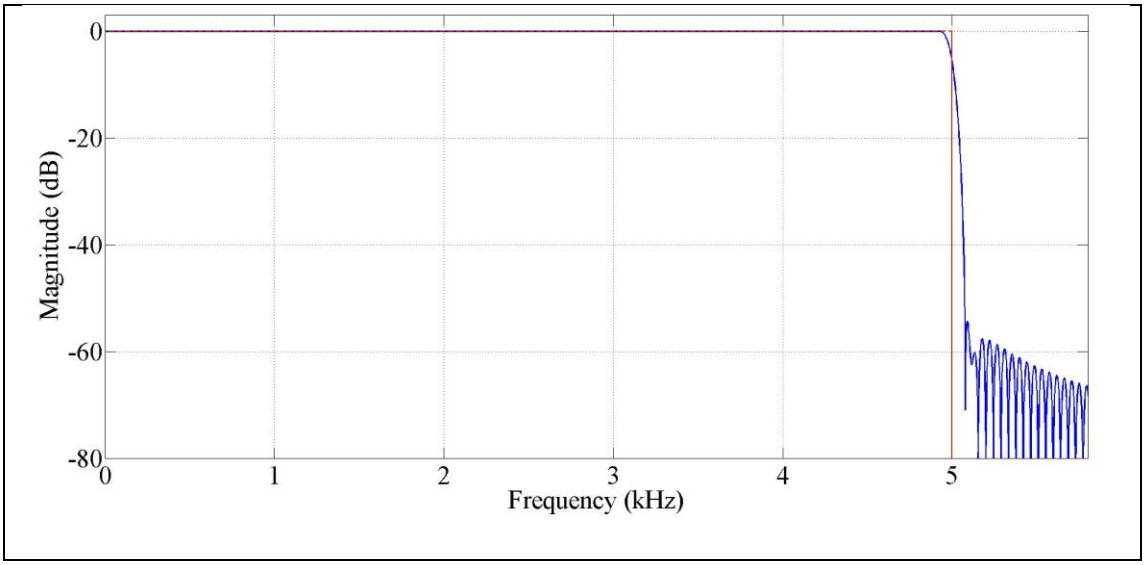


Figure 30 – Upper panel: impulse response of a low pass filter. Bottom panel: Frequency response of a low pass filter. The red line indicates its ideal frequency response.

The stimulus used for the measurements was randomly generated (by permutation of the pitch parameter). Presto-chirps were played randomly either ascending or descending, and all were 0.125 seconds long. Presto-chirps were spaced apart by 0.1 seconds long silence, which was also the length of the used analysis window. The total length of the stimulus was of about 18 seconds, as it can be seen in the graphical time-frequency representation shown in Figure 31. The frequency axis is in logarithmic scale.

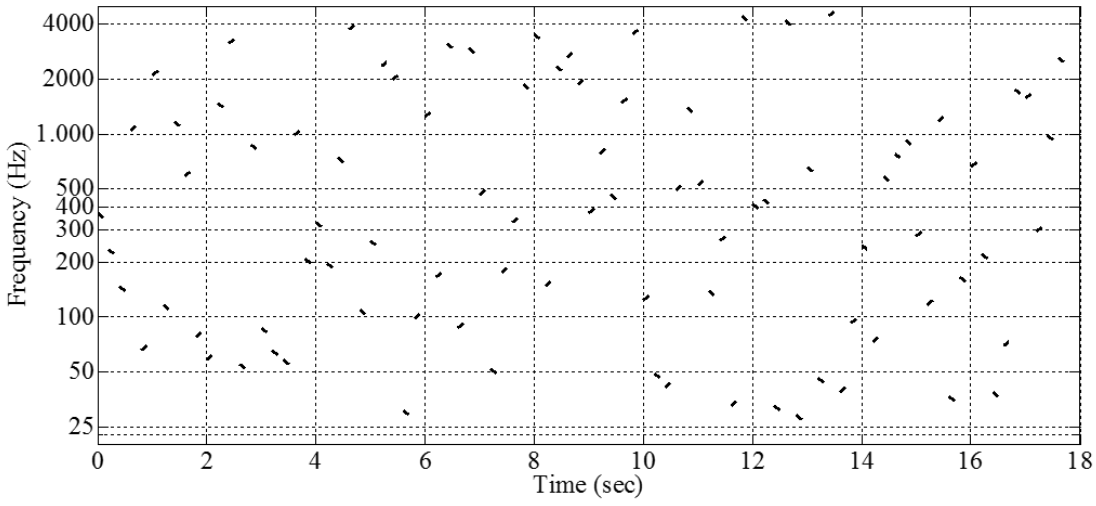
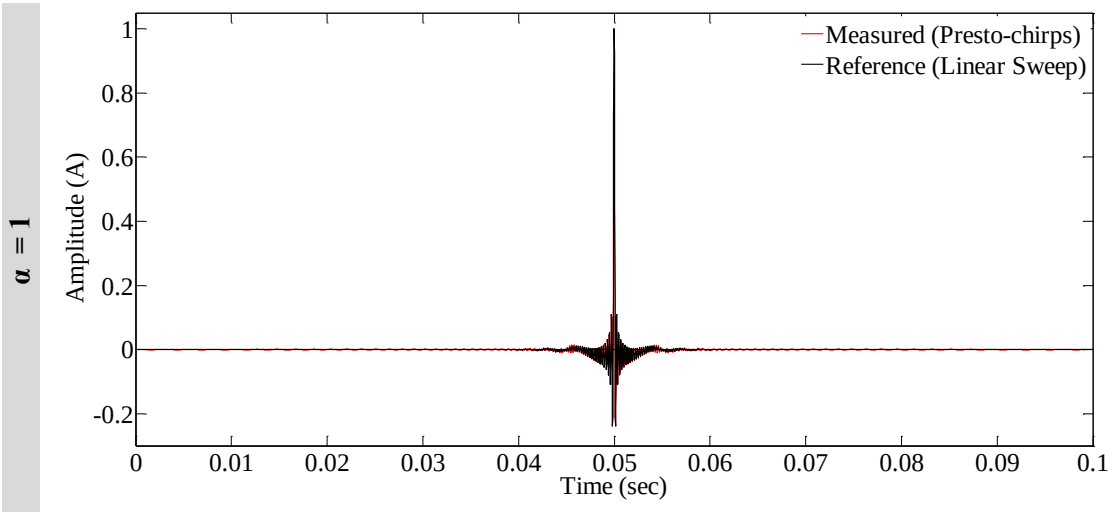


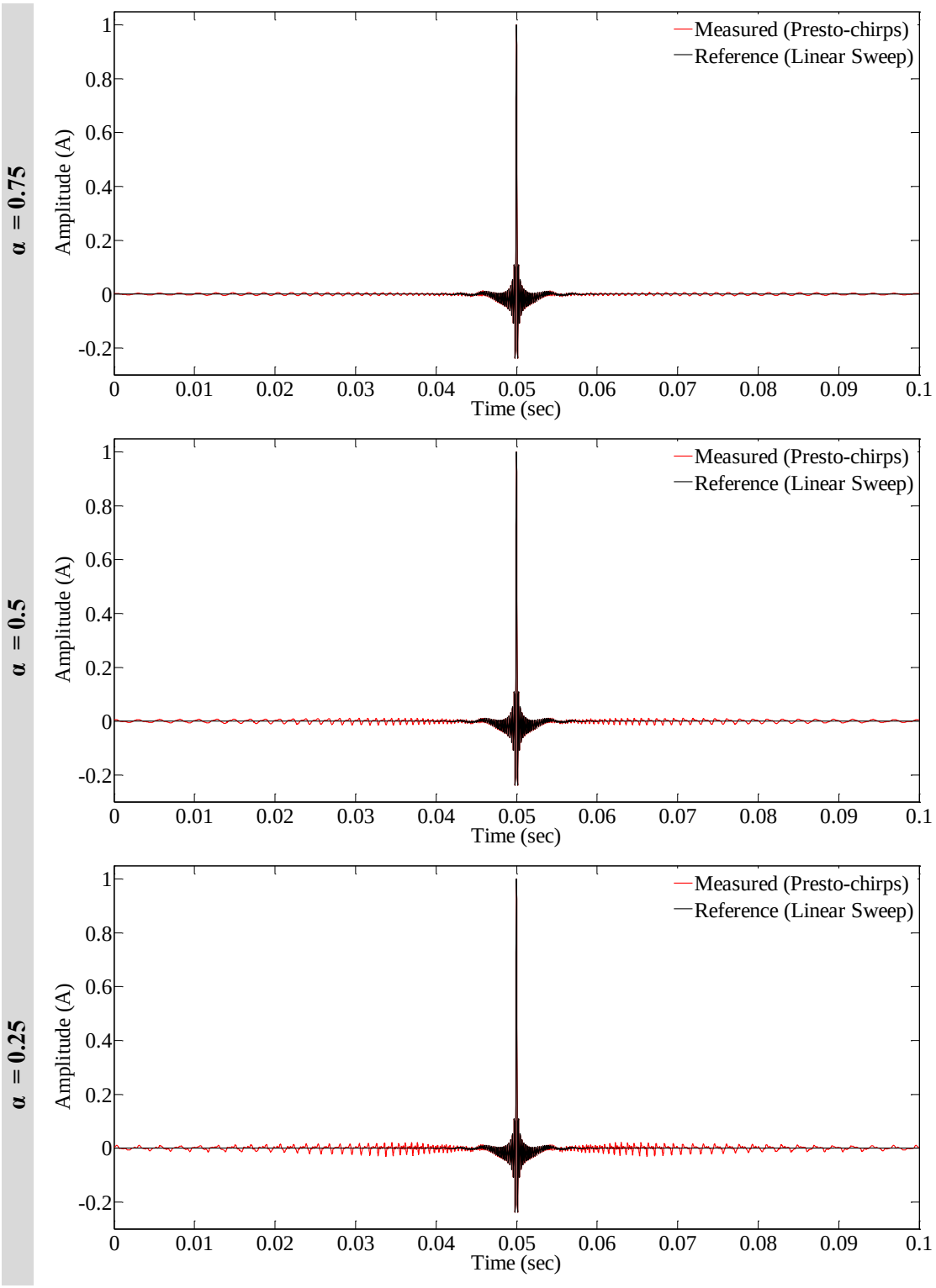
Figure 31 – Time and Log Frequency graphic representation of the stimulus.

By the point of view of music, 0.125 seconds correspond to a semiquaver (note) of $1/8^{th}$ of a second for a music time of 120 beats-per-minute (BPM). This duration is suggested has a sensible "minimum" duration for presto-chirps. Shorter lengths are possible but have not been experimented in this thesis. The decision of using a minimum duration has been taken based on experimental evidences, which showed that short chirps produce an increased spectral leakage compared with longer duration. This issue is discussed in Section 4.2.1.5. Another reason is for the low energy that a short chirp will deliver into the tested system. Analytical and experimental studies of this issue are proposed as further work.

4.2.1.1 Time Domain Analysis

Figure 32 shows the comparisons between the impulse response of the low pass filter (the reference IR) and the IRs measured (in red) using the presto-chirps stimuli windowed as for the α values labelled in the left (grey column). The measured IRs show the presence of a "spiky ringing" bias, which becomes more pronounced as α decreases.




α = 0.75
α = 0.5
α = 0.25
Measured (Presto-chirps)
Reference (Linear Sweep)
Amplitude (A)
Time (sec)
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1

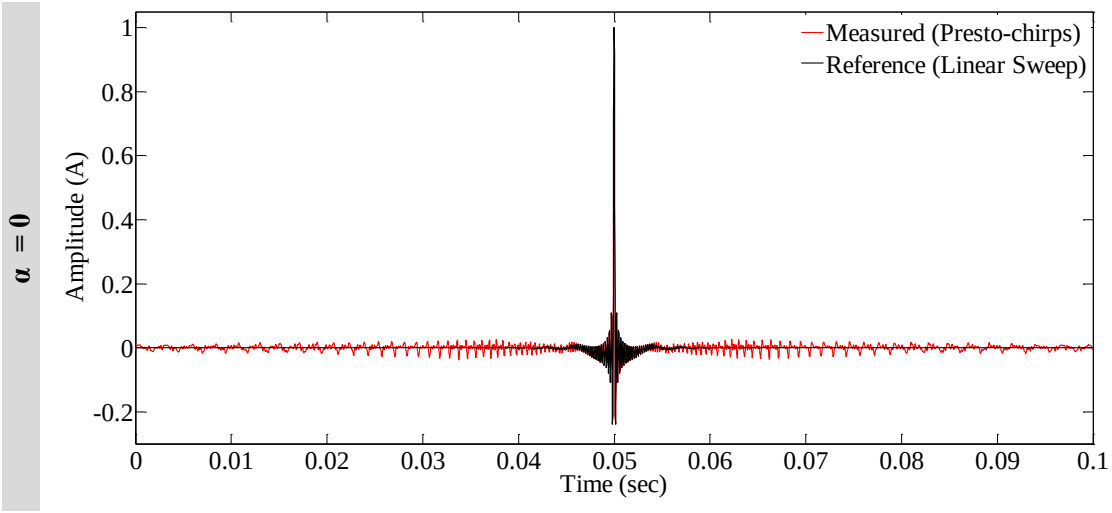


Figure 32 - Amplitude differences functions for several α values.

4.2.1.2 Synthesis Noise

The bias introduced in the measured IRs, which will be hereafter refereed as the "synthesis noise" (because it is caused by the "non-perfect" synthesis of the partial impulse responses) is the time domain counterpart of the frequency bias discussed in Section 3.4. It is, in practice, the autocorrelation functions of the used stimuli. The distribution and the amplitude of the ringing spikes are on one hand, related to the length of each presto-chirp composing a stimulus, on the other hand, to the specific window applied to them.

To shed a light on the relation between duration and accuracy, in the following will be compared measurements performed using the stimulus presented in Figure 31 (0.125 seconds-long presto-chirps), with measurements performed the same stimulus, but formed of one second-long presto-chirps.

For example, Figure 33 shows the visual comparison between the reference impulse response of the filter (in black), and the measured impulse responses of the same filter

(in red) calculated using 0.125 seconds long (upper panel), or 1 second long (bottom panel), rectangular-windowed presto-chirps. The IRs are shown magnified to enhance the visualisation of the noise; the analysis window has been extended to one second to increase the resolution (the presto-chirps have been spaced accordingly). In the former case, the measured IR presents higher side lobes and a sparse ringing compared with the latter case.

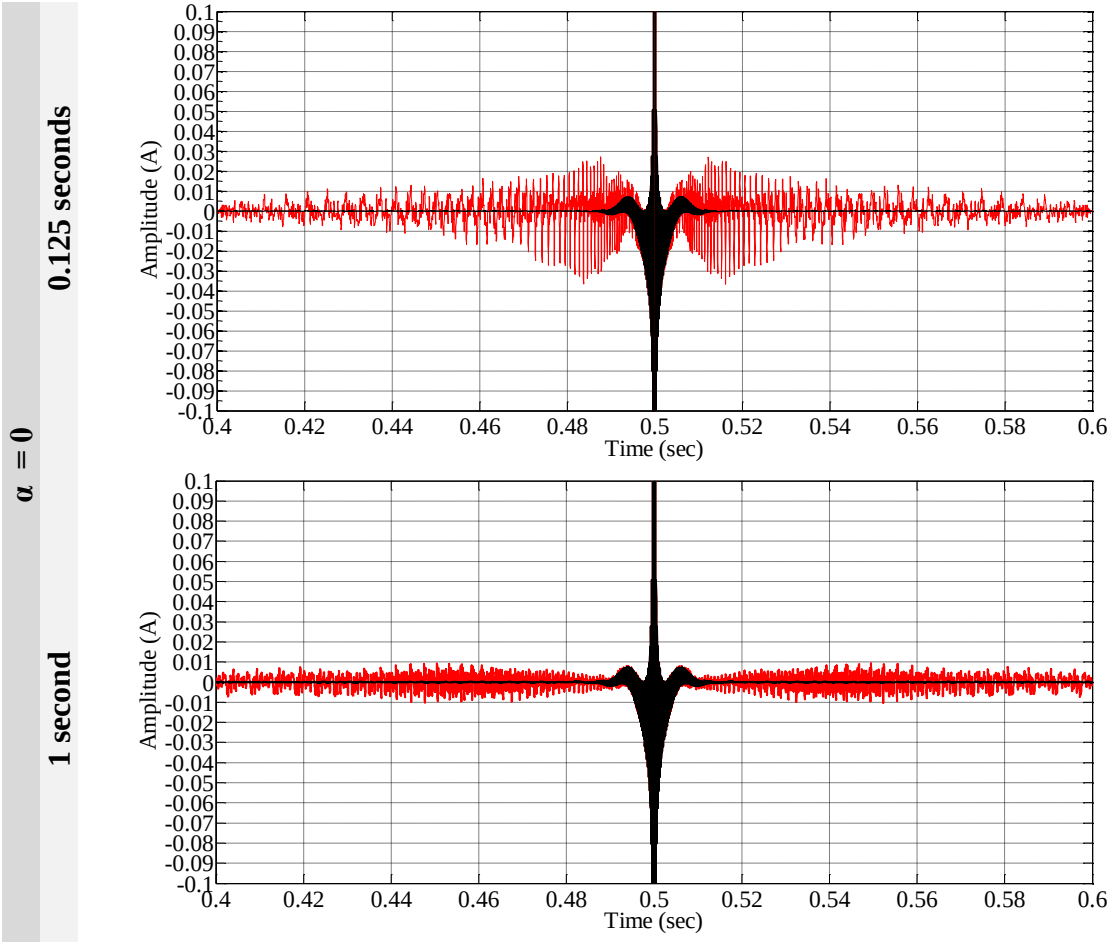


Figure 33 – In black is represented the reference IR (a low pass filter). In red is represented the IR measured using rectangular-windowed presto-chirps with duration of 0.125 seconds in the upper panel, and of 1 second in the lower panel.

On the contrary, Figure 34 shows uses the same durations for the presto-chirps used in Figure 33, but Hanning-windowed. In this case, besides the side lobes have lower

amplitude compared with the previous case, their change only slightly for the two different durations tested. The one-second long presto-chirps have lower side lobes but longer ringing, compared with the 0.125-seconds long presto-chirps.

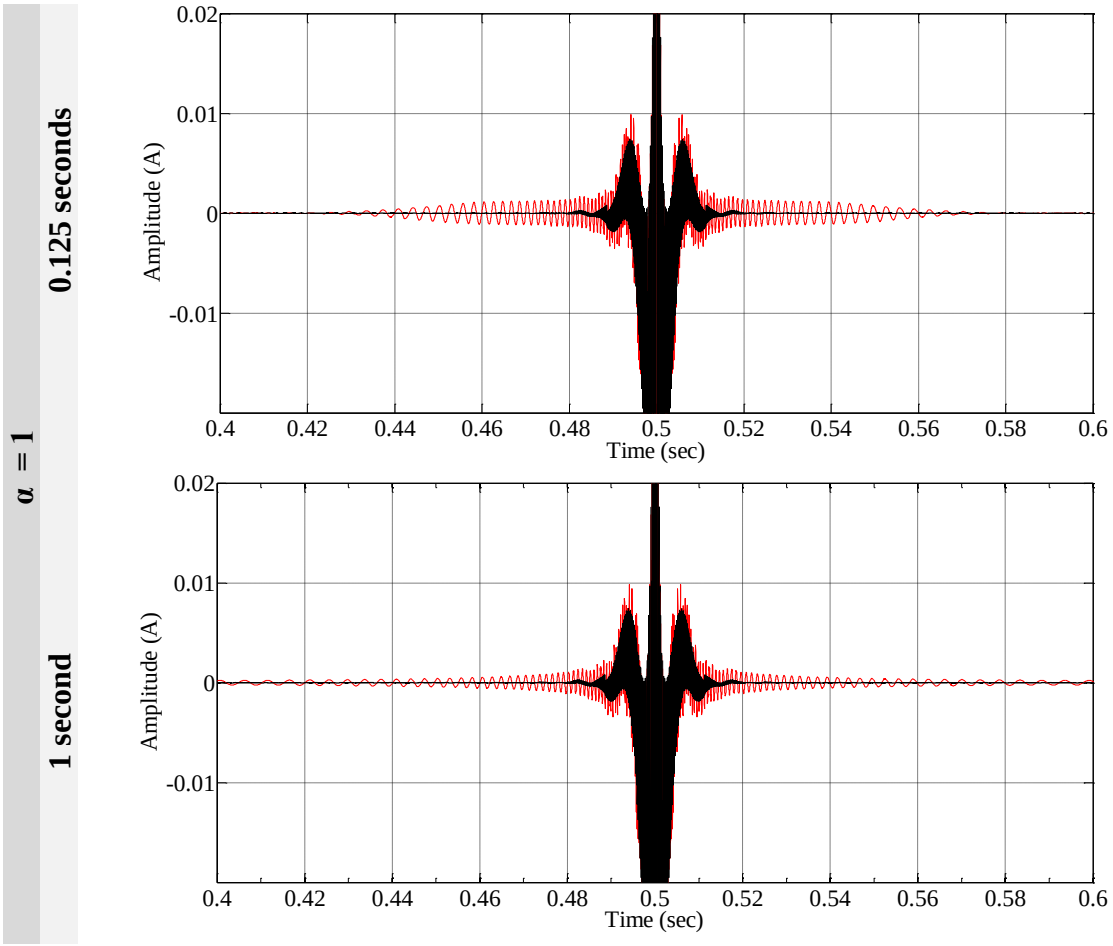


Figure 34 – In black is represented the reference IR (a low pass filter). In red is represented the IR measured using Hanning-windowed presto-chirps with duration of 0.125 seconds in the upper panel, and of 1 second in the lower panel.

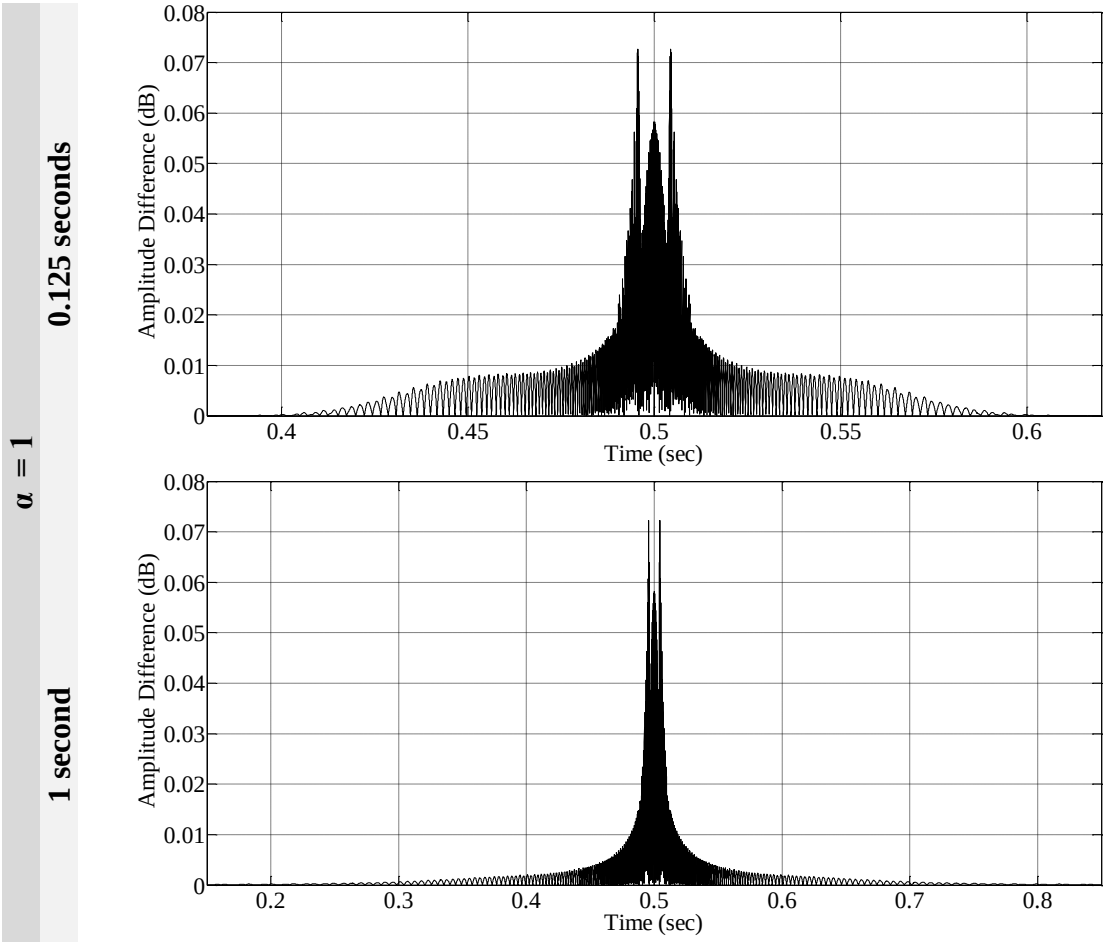
In general, for both cases, the use of shorter presto-chirps resulted in IRs having shorter ringing effect but higher side lobes, and vice versa. This effect by the point of view of filter theory (presto-chirps are similar to pass band FIR filters) is because longer presto-chirps have steeper transition bands. Whereas, shorter presto-chirps results in filters having smoother transition bands (at the same time they may be less

close to the target/ideal frequency response). Since the ACFs of the stimuli are related to the duration of the presto-chirps, this implies that for stimuli composed using presto-chirps having several different durations, the resulting ACFs will be the combination of all the autocorrelation functions of every single presto-chirp composing the stimuli. In measuring RIRs, since the ACFs of the stimuli is theoretically convolved with each peak of the RIR, the side lobes of the ACFs might interact, constructively or destructively, with the neighbouring peaks, resulting in what has been afore defined "synthesis noise".

4.2.1.3 Amplitude Differences

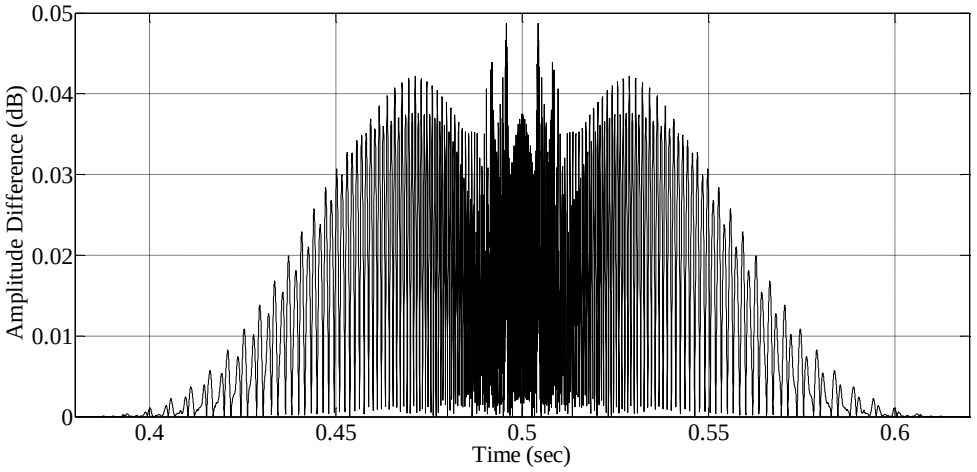
Figure 35 shows the amplitude differences between the reference IR, and the IRs measured using presto-chirps stimuli Tukey-windowed for its α parameter equal to the value labelled in the left. The amplitude differences are calculated, for each window, for two cases. In the upper figures, the presto-chirps composing the stimulus had duration of 0.125 seconds. In the lower figures, they had duration of 1 second. These analyses report the relation between the ACFs of the stimulus, and the duration of presto-chirps by which it is composed. The two tested durations represent the minimum and maximum durations, for the presto-chirps, used in this thesis. The use of short presto-chirps results in ACFs of the stimuli with higher side lobes compared with the use of longer presto-chirps. This suggests that the use of longer presto-chirps can lead to a greater accuracy, compared with the use of shorter presto-chirps. In addition, for the Hanning-windowed case the ACFs have lower side lobes. It must be noted though, that the amplitude differences of the central points (at 0.5 seconds), which correspond to the main peak of the measured IRs, do not vary much for the two different durations for the same case. However, it varies when different windows are

used; being slightly higher for the Hanning window case compared with the other cases. These results may be interpreted as the inherent error that using the presto-chirps method will add the each peak of a measured RIR (only for stimuli composed of fixed duration presto-chirps, for the two tested durations).

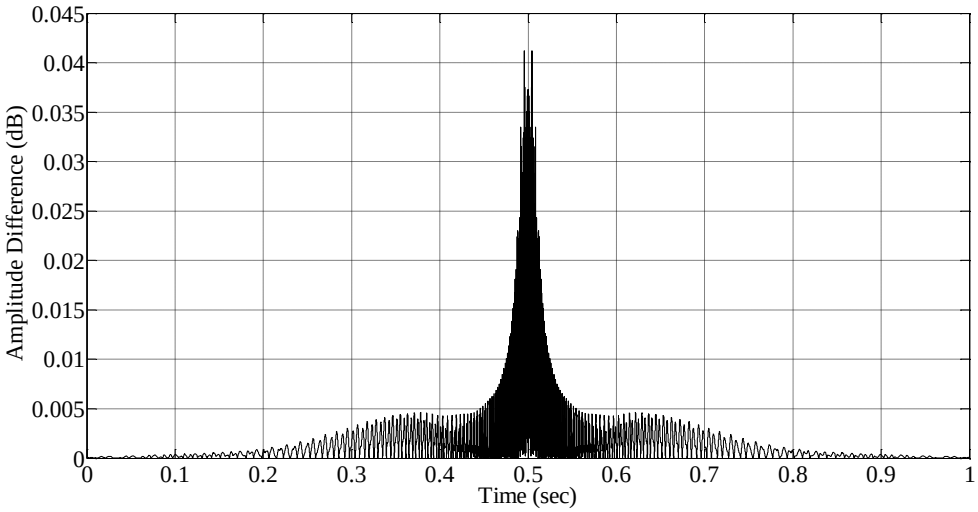


α = 0.75

0.125 seconds

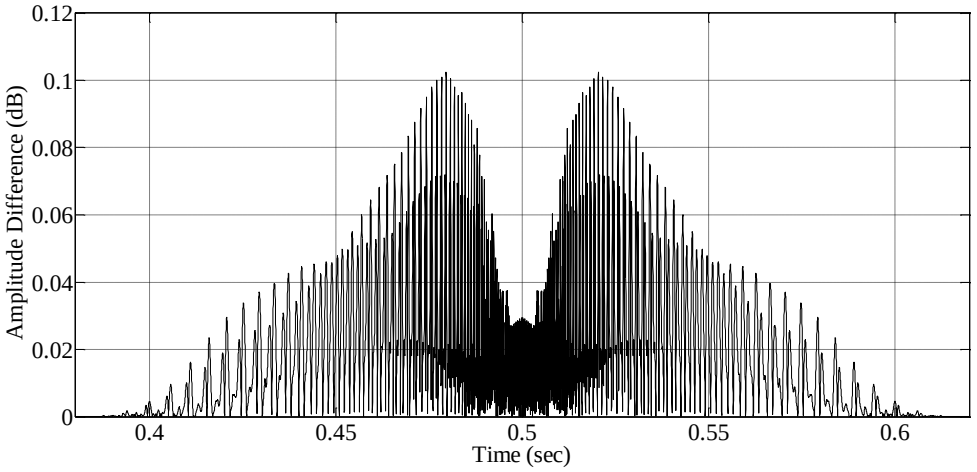


1 second

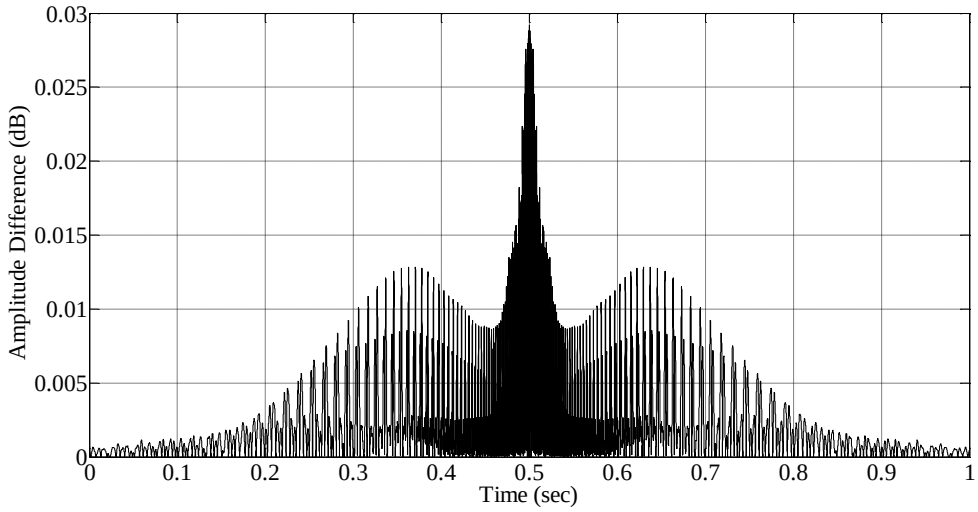


α = 0.5

0.125 seconds

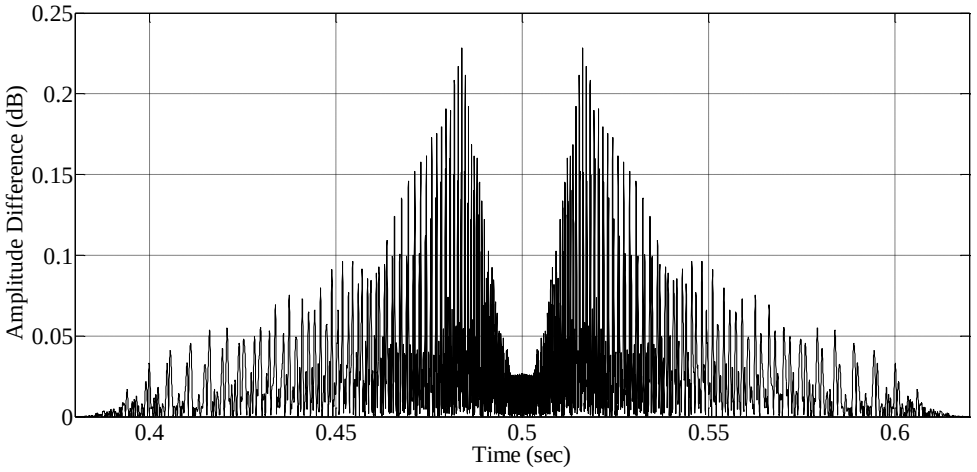


1 second

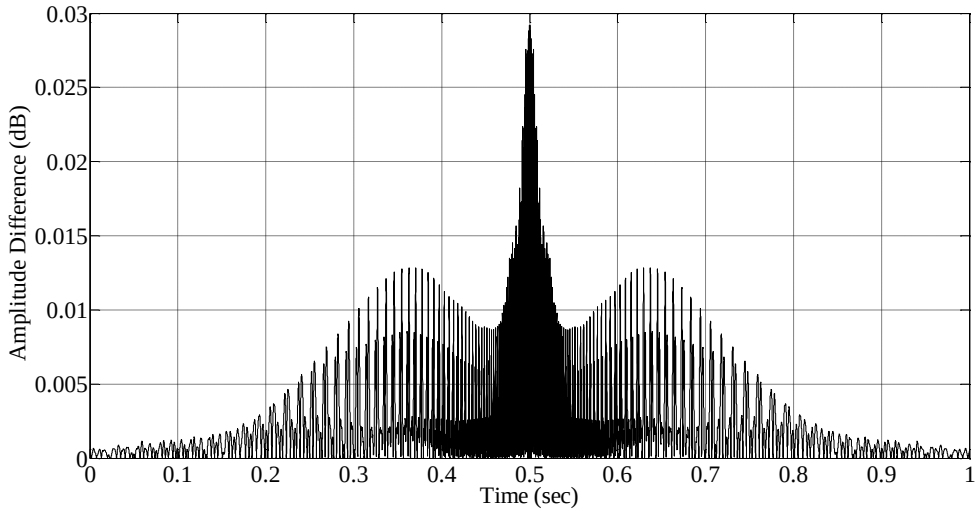


α = 0.25

0.125 seconds



1 second



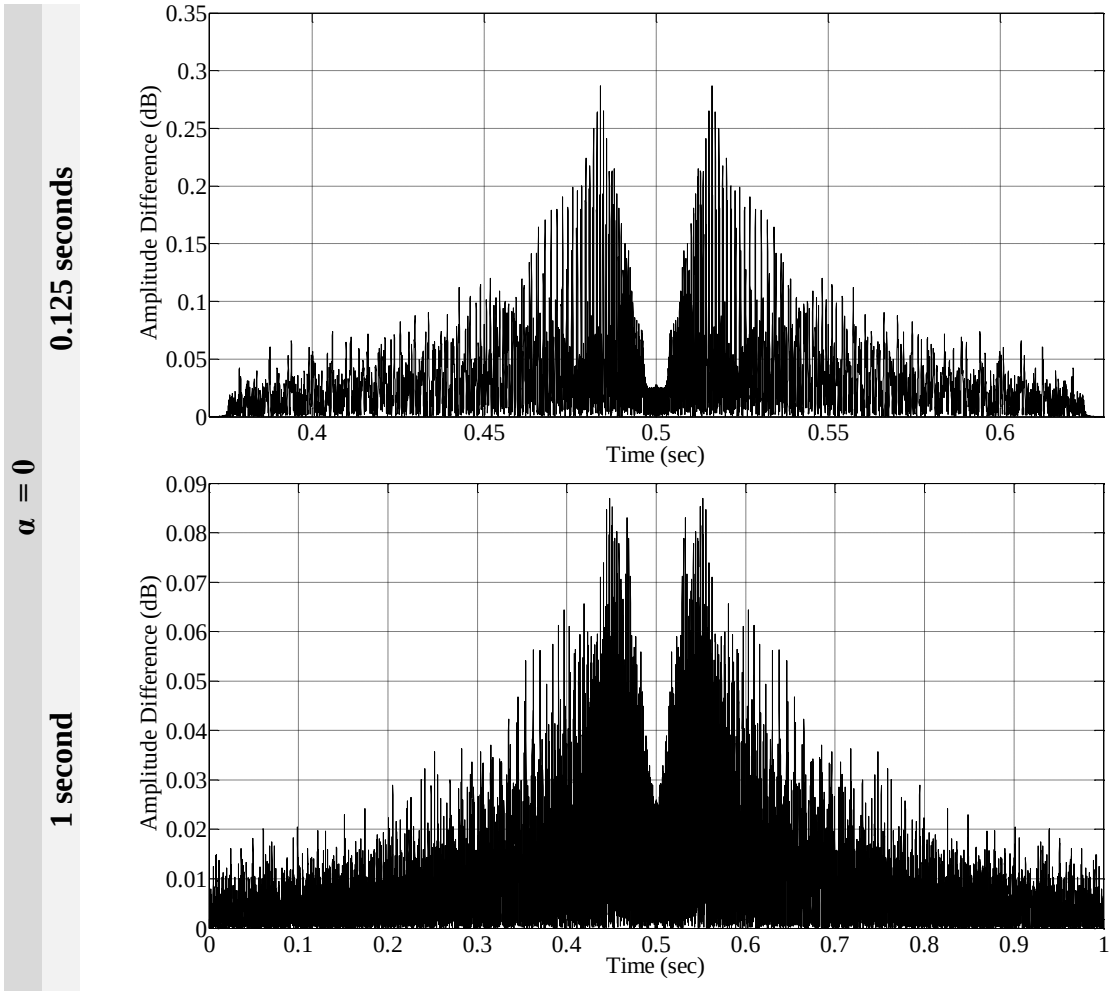


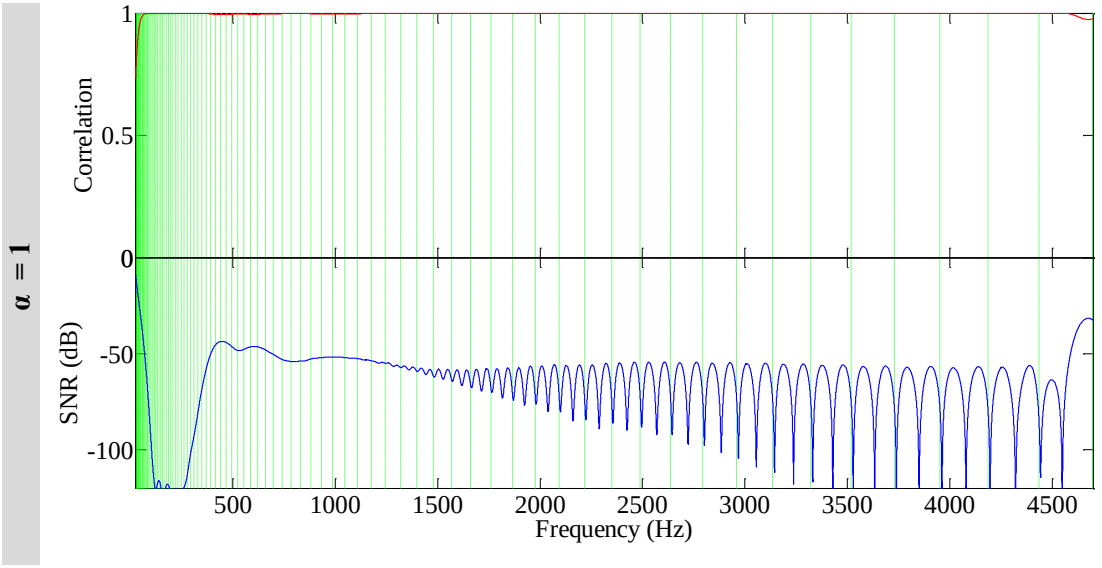
Figure 35 - Amplitude differences versus α.

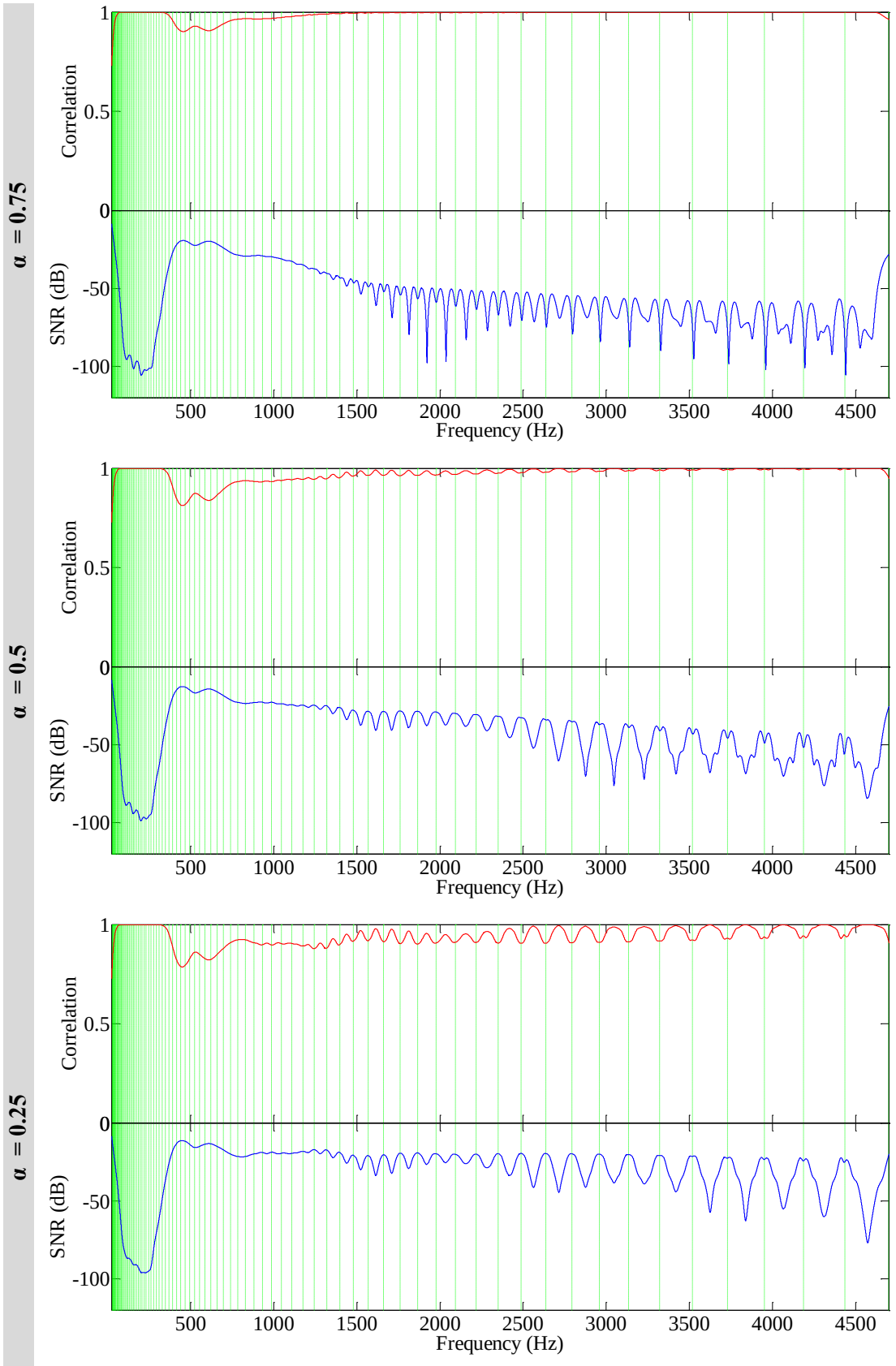
The results shown in this section are related to the Tukey window features that have been discussed in Section 3.1.1, which are consequently related to the behaviour/theory of window functions. In general, smoother windows have lower side lobes but larger main lobe (which comports larger estimation errors), and vice versa. However, for the aim of calculating RIRs the optimal solution is achieved by using a window that has lower side lobes. This, to limit the interaction of the synthesis noise between the peaks of a RIR. From the obtained results, the use of a Hanning window is suggested as the solution that will introduce the minor amount of error in the measured RIRs. This statement will be further confirmed from the results achieved in the others validations presented in this thesis.

4.2.1.4 Frequency Domain Analysis

Figure 36 reports the coherence and SNR functions between the measured and reference RIRs. For α=1 the coherence is one for almost all the measured spectrum. For the use of Tukey windows with α<1, there is the bias increase progressively as α lowers. This consequently decreases the coherence. The SNR resembles the shapes of the presto-chirps, with local minima at those frequencies equal to the frequencies of musical notes (signalled by the dashed vertical lines).

An interesting effect that has to be noted is that, although both the coherence and the SNR worsen towards lower frequencies, they result maximised below about 350 Hz, for all the tested configurations. The frequencies below 30 Hz are interested by and "edge" differences between the FRFs, due to slightly different transition bands (the reference IR had slightly more smoothed frequency boundaries).




α = 0.75
α = 0.5
α = 0.25
Correlation
SNR (dB)
Frequency (Hz)
1
0.5
0
-50
-100
500
1000
1500
2000
2500
3000
3500
4000
4500

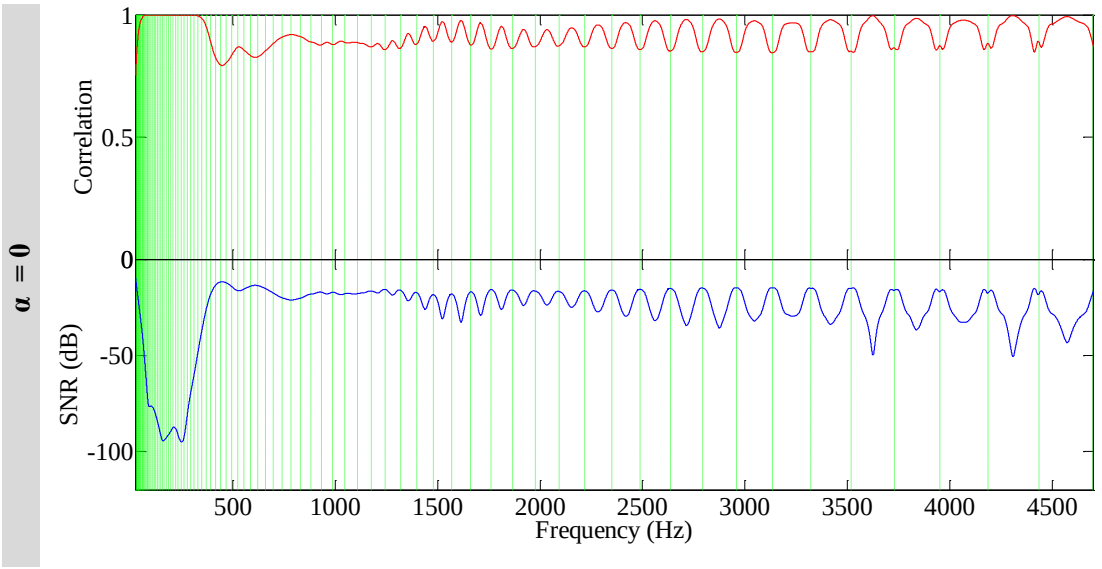


Figure 36 – Coherences function as a function of α.

4.2.1.5 Spectral Leakage

The trend of the SNR for frequencies below 350 Hz can be explained by considering that the bandwidth of a signal is related to its duration. Presto-chirps are conceptually similar to linear-phase FIR filters, therefore their shapes (thus their magnitude response) is also determined by their lengths (in samples), which correspond to the taps of a FIR filter. For example, a second long presto-chirp sampled at 44.1 kHz is conceptually similar to a 44099 taps FIR filter. Note that for FIR filters theory an odd number of samples is needed to achieve a linear phase filter, since a symmetrical IR is needed.

Figure 37 shows the spectra of a Hanning-windowed presto-chirp, for the durations reported in the legend. It can be seen that longer presto-chirps presents wider transition bands. A widening of the transition bands results in a greater overlapped area between adjacent presto-chirps, as it can be seen in Figure 38, which reports the autospectra of (a part of) the set of Hanning-windowed presto-chirps with duration

varying form 0.125 seconds up to 2 seconds. The autospectra of presto chirps are the channels of the filter bank used to analyse impulse responses.

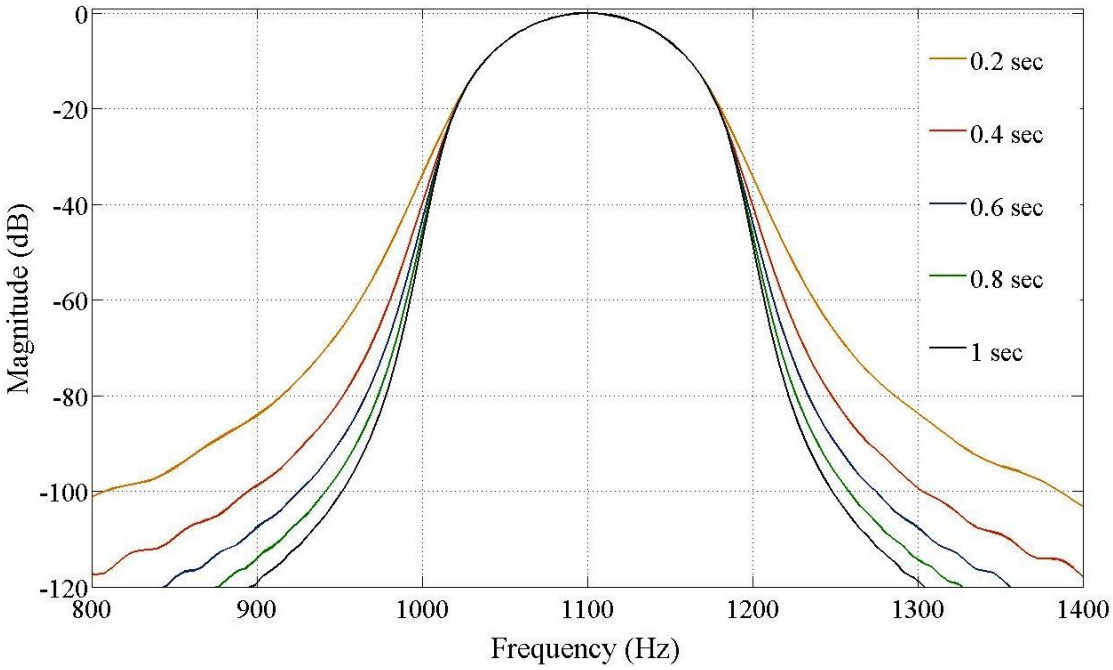
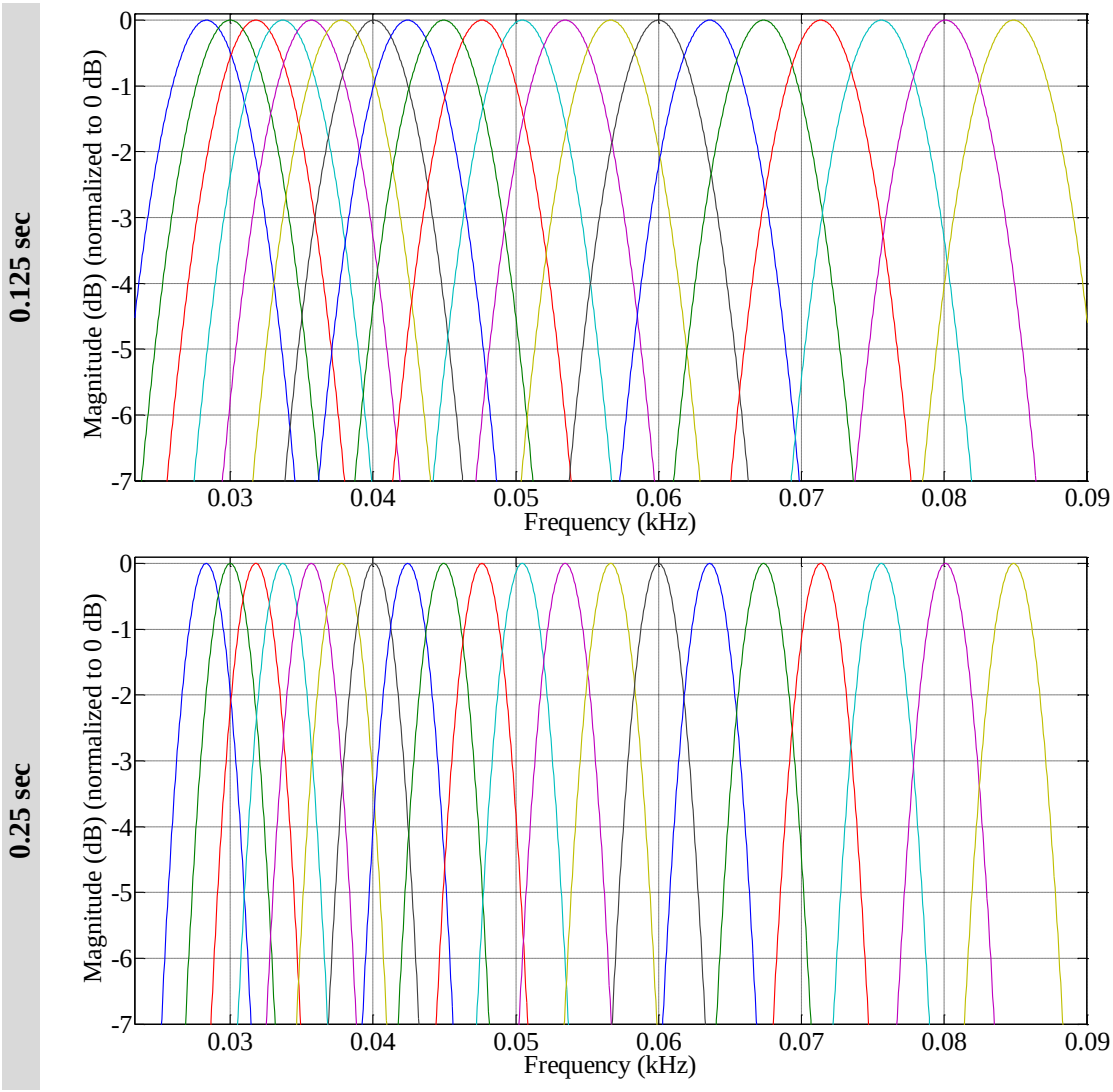


Figure 37 – Presto-chirp frequency shapes versus their durations (α = 1).

It can be noted that for the duration of 0.125 seconds, the first two autospectra (leftmost part of the filter bank) overlaps at about 0 dB. Conversely, the last two (shown) autospectra overlaps at about 1 dB. For longer presto-chirps durations, the overlapping points between adjacent autospectra lowers, stabilising progressively, starting from the higher sub bands, at about -6 dB. In practice, the narrower the bandwidth of presto-chirps the greater the number of taps (thus a longer duration) needed to reach a stable frequency response. Whereas, wider bandwidths reach a stable spectral configuration yet from shorter durations. Figure 36 shows the (wide) overlapping between the autospectra for presto-chirps with duration of 0.125 seconds. This could be reason for the high SNR at frequencies below 300 Hz. Thus, since the low frequencies are excited/analysed by several presto-chirps (with more or less

power), this might produce an averaging effect that consequently increases the SNR. In other words, the cumulative power of a set of presto-chirps results in more energy being fed at low frequency bands rather than at high frequency bands. This relates to the spectral distribution of the sets of presto-chirps shown in Section 3.2.



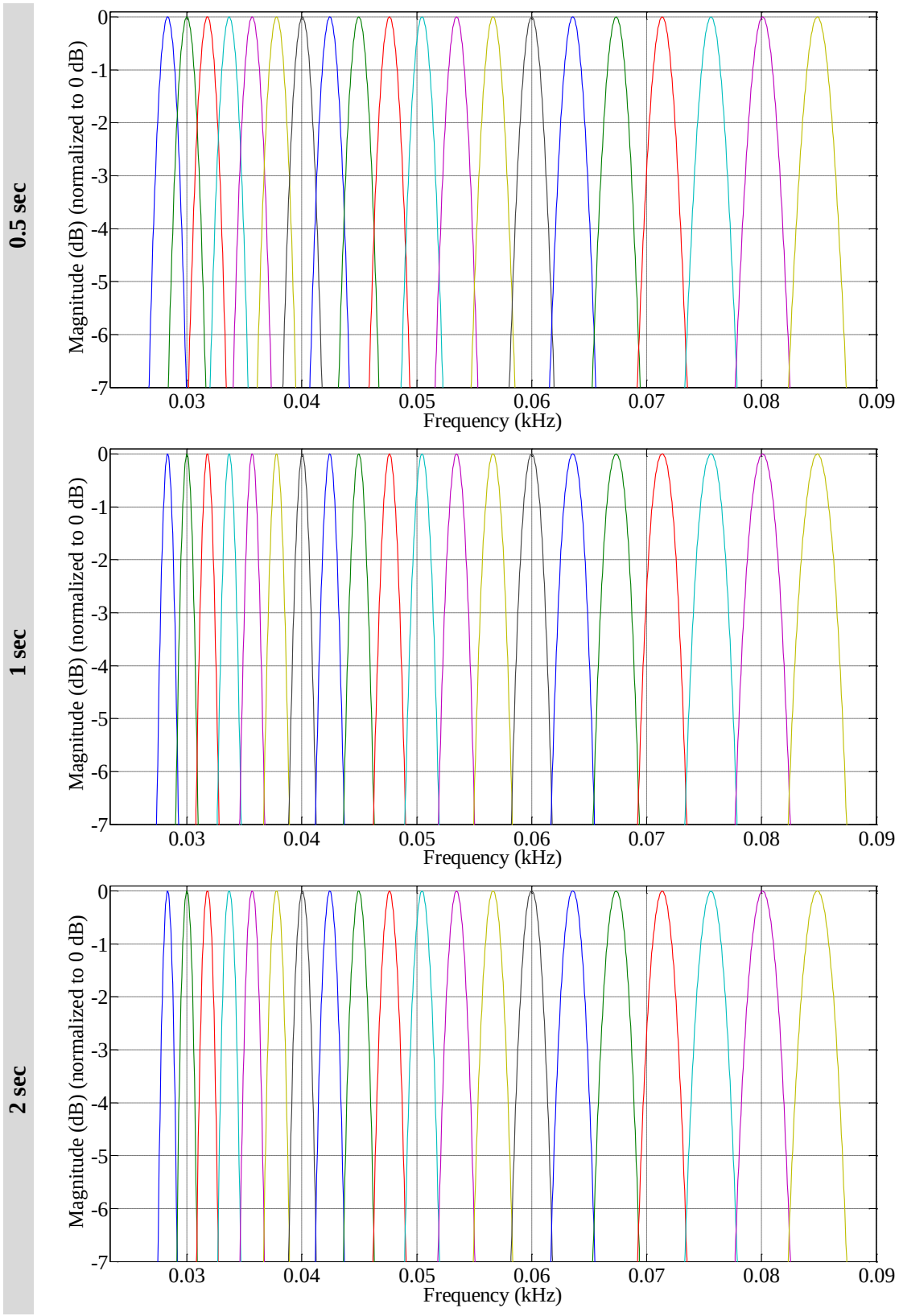


Figure 38 – Autospectra of the beginning part of the set of presto-chirps versus presto-chirps' duration. For longer duration that transition bands narrow till a "stable" configuration, for duration = 2 seconds.

To further prove this deduction, Figure 39 reports the coherence and SNR analyses of a measurement performed using a stimulus composed by one second long Hanning-windowed presto-chirps. The results is that the "averaging effect" is shifted to the left end of the spectrum, compared with what Figure 36 shows. The increased overlapped area between adjacent presto-chirps is the reason because in Section 3.6 it has been suggested to use a "semitone" distance larger than three semitones, for the low frequency notes, when they have duration lower than one second.

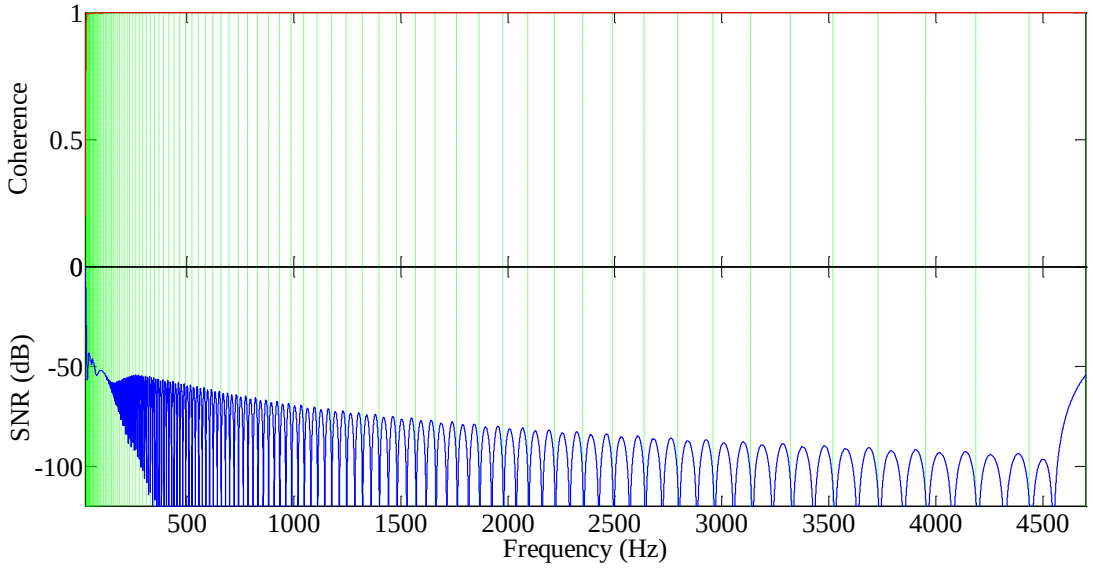
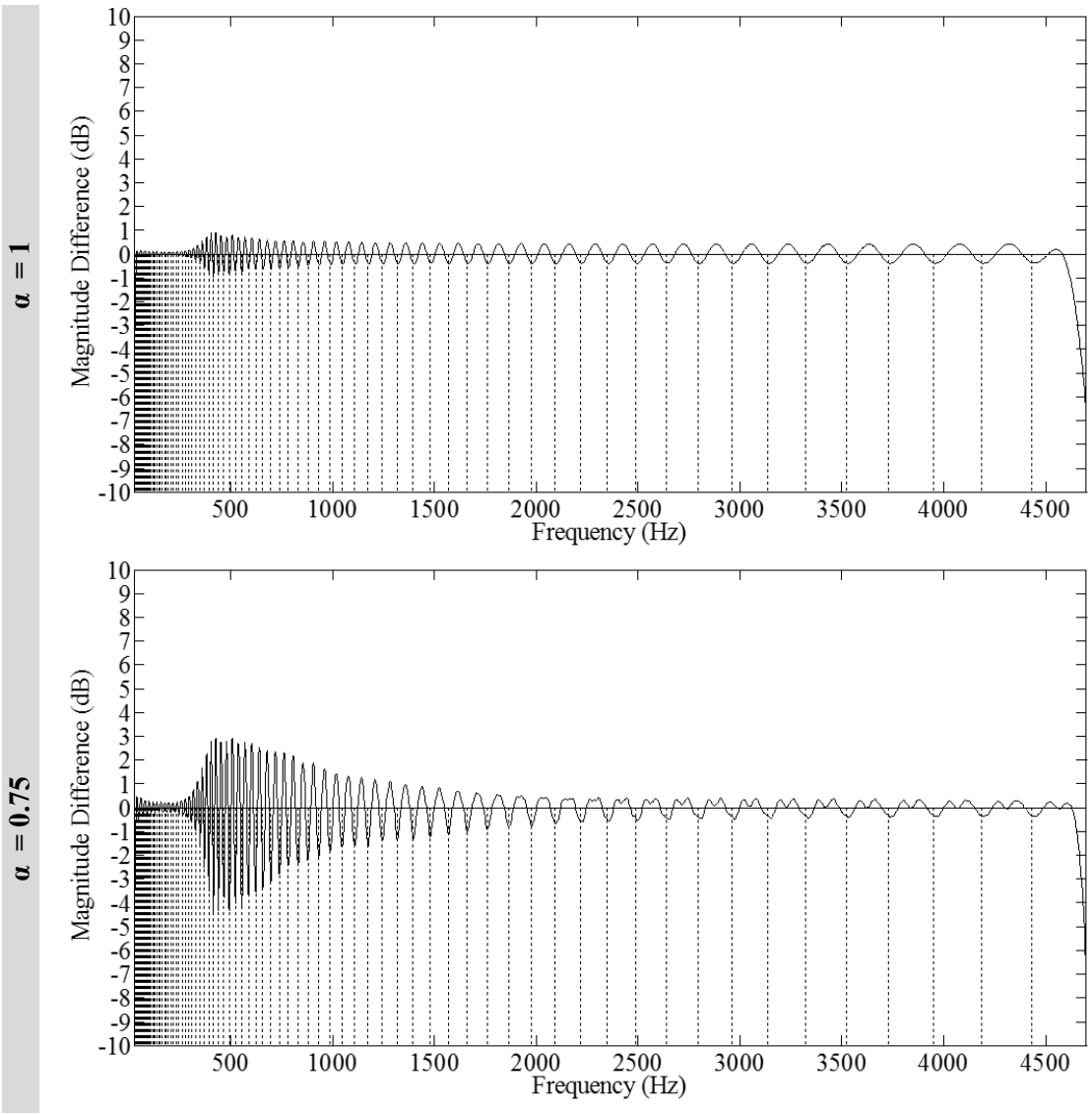


Figure 39 - One second long PCs – alpha = 1.

4.2.1.6 Frequency Differences and Averaging Effect

Frequency differences, between the reference and the measured IRs, are calculated using a stimulus compose of 0.125 seconds-long presto-chirp, as for Eq. (59). As Figure 40 reveals, the use of a Hanning window gives the "flattest" spectrum (less introduced bias) compared with the others spectra, which instead present a progressively increasing bias as α decreases. For all configurations the reductions of the frequency differences at the low end of the spectrum, due to the averaging effect,

is clearly visible. To further extend the discussion, Figure 41 reports the magnitude differences (in the range 25 Hz – 500 Hz, for a clearer visualisation of the phenomenon), between the reference RIR and the RIRs measured using Hanning windowed presto-chirps, for the labelled durations of 1/8, 1/4, 1/2, 1, and 2 seconds. These durations were chosen because commonly used for the duration of notes in music (BPM = 120).



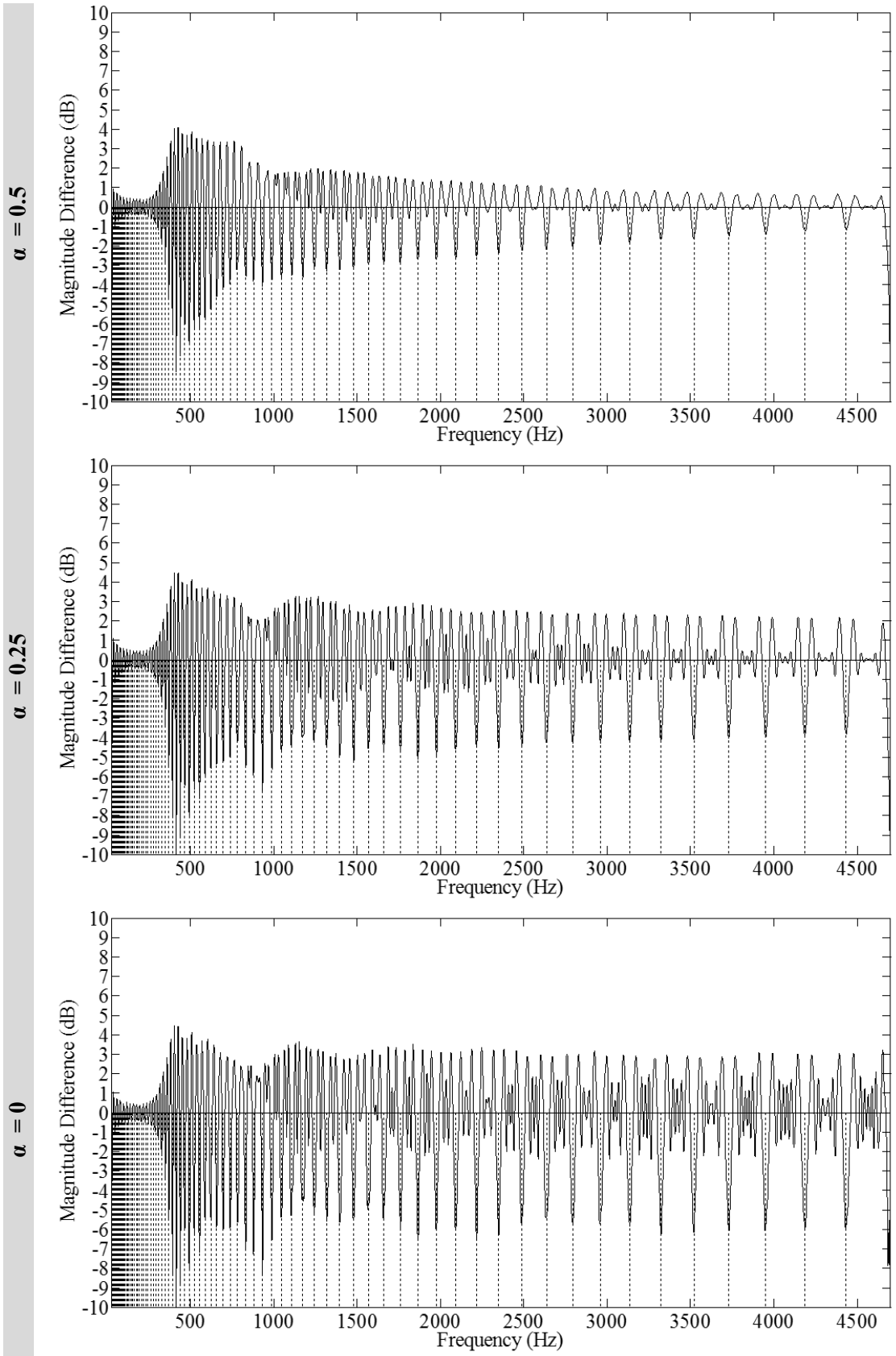


Figure 40 - Magnitude differences for several α values.

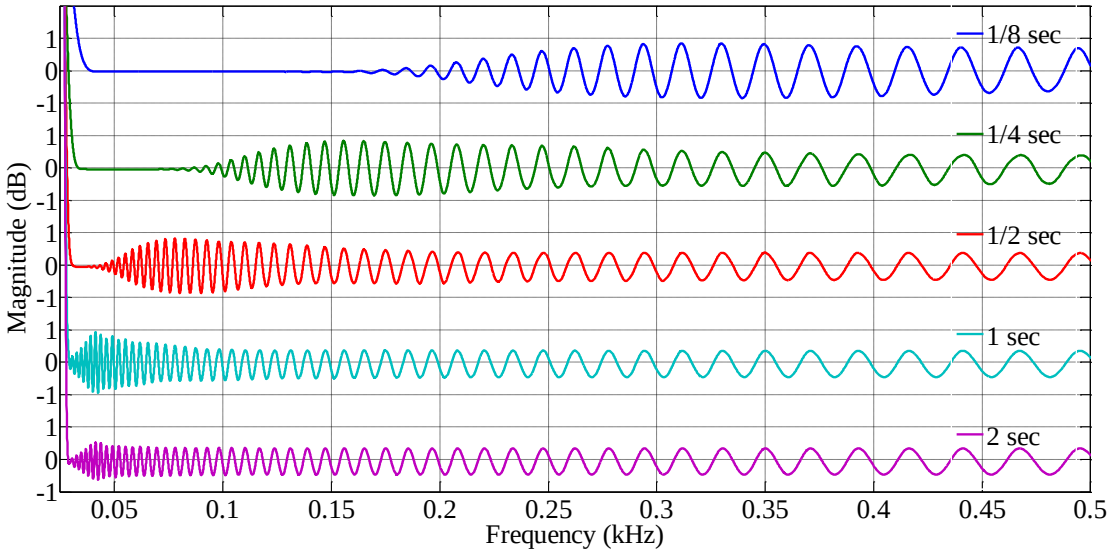


Figure 41 – Magnitude differences functions for a stimulus composed of Hanning windowed presto-chirps (each) having the durations specified in the figure.

As the duration of the presto-chirps increases, the frequency differences tend to have higher values towards the low end of the spectrum. This effect is caused by the different amount of overlapped area between presto-chirps, as it has been shown in Figure 38. The higher magnitude difference applies to all of the durations, but for the 2 seconds, durations bounded, durations bounded at about ± 1 dB, apart from the 2 seconds. This is the error that is generally introduced by measurements performed using a Hanning windowed presto-chirps stimulus in the frequency domain of measured RIRs. However, as it can be seen, for presto-chirps with duration of 2 seconds the introduced bias becomes of ± 0.5 dB. For longer durations of 3, 4, and 5 seconds, reported in Figure 42, the bias does not vary. This highlights that the presto-chirps frequency response is stable (thus increasing further the number of taps does not produce sensible differences in the magnitude response of a filter). However, although those latter lengths might not be of a practical use for presto-chirps measurements, this result poses the question of whether using higher sampling

frequencies might produce similar frequency responses also for presto-chirps that also have a shorter duration.

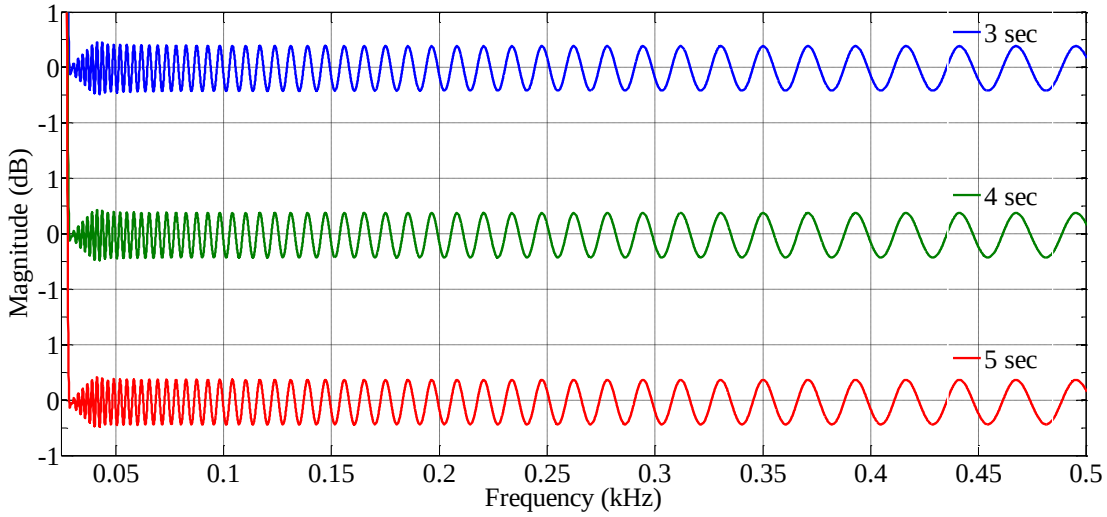


Figure 42 – Magnitude differences functions for a stimulus composed of Hanning windowed presto-chirps each having the durations specified in the figure.

4.2.1.7 Discussion on the Case Study

This case study has been useful to highlight, from different points of view, the bias that presto-chirps stimuli introduce in measurements. For the tested configurations have resulted that using Hanning windowed presto-chirps gives the best accuracy. This is also confirmed in the correlation analysis, reported in the graph in Figure 43, which shows the degree of correlation between the (whole) measured RIRs and the reference one as a function of α. The lower α, the less the RIRs are correlated with the reference, which in turn denotes a greater amount of introduced synthesis noise in the measurements. It has also been highlighted that the duration of presto-chirps has direct consequences on the synthesised frequency response of the stimulus. In this regard, it has been shown that the spectra of presto-chirps become wider as their duration decreases, especially at low frequencies. This increases the overlapped area

between neighbouring presto-chirps, which on one hand results on an averaging effect that can increase the SNR, and on the other hand might slightly complicate composition arrangements.

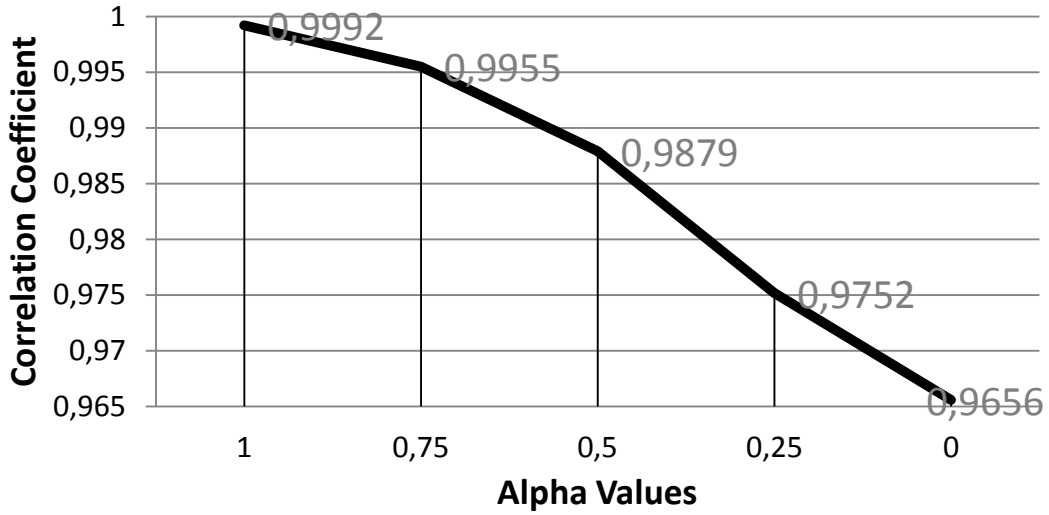
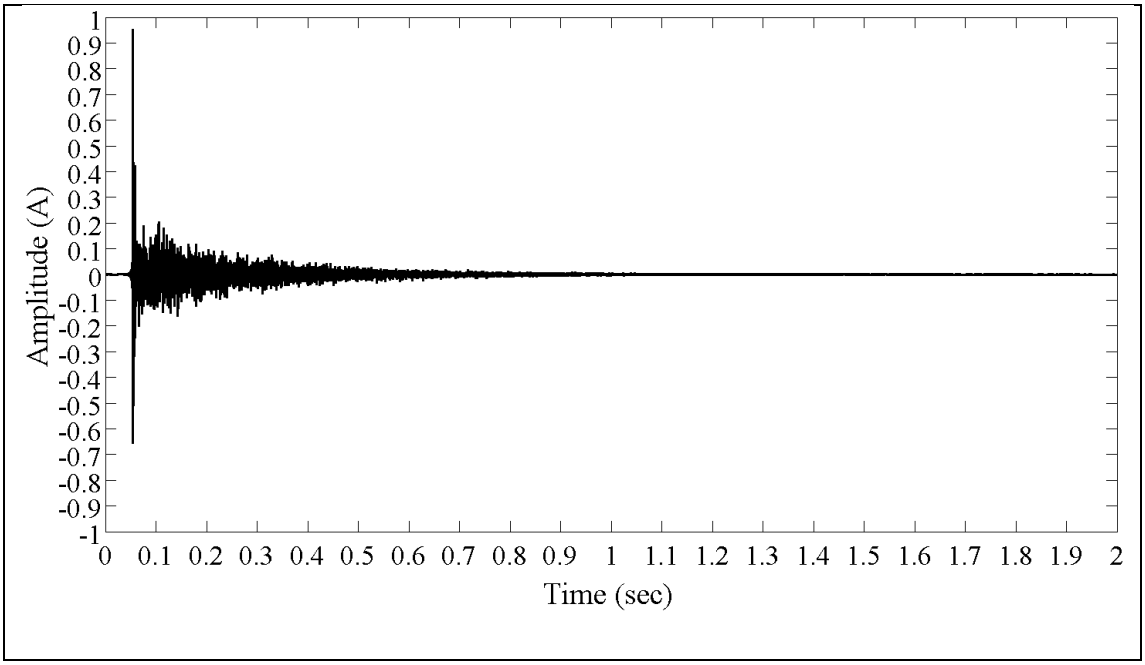


Figure 43 – Correlation analysis between the whole reference RIR and the measured ones, as a function of the window configuration used.

4.2.2 Case Study: St. Patrick Church

In this case study the recorder RIR of a church, downloaded (with permission) from the site OpenAirLib.net (http://www.openairlib.net/) is used as the reference RIR. Reported from the site: "*The St Patrick's Church in Parrington (U.K.) dates back to the medieval era and expresses the decorated design of English Gothic Churches of the early 14th century. It has a cruciform plan, with an external length of about 46 meters, a breadth of 27 meters and a height of 57.5 meters up to the very top of the central tower*". Its RIR, as reported, was calculated using a 15-second long ESS (fed by a Genelec S30D loudspeaker, and measured using a Neumann KM 140 microphone) ranging from 20 Hz up to 22 kHz. The RIR, for the purpose of comparisons, has been band pass filtered in the frequency range 27 Hz - 4.7 kHz. Its impulse and frequency responses are shown in the upper and bottom panel of Figure 44, respectively. The averaged RT60 is of 1.85 seconds.



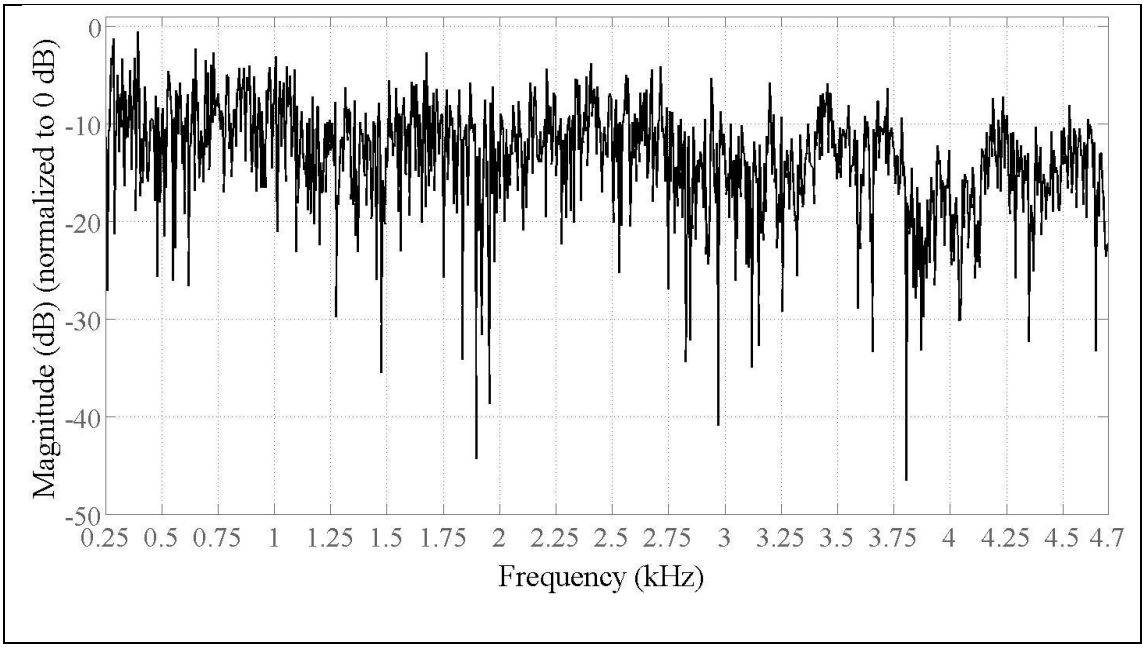


Figure 44 – In the upper and bottom panel are shown the impulse and frequency response of the tested RIR, respectively. Band limited in the range 27.5 Hz – 4.69 kHz.

Figure 45 shows the 9 seconds long stimulus used in this simulated case scenario. The stimulus is composed of three "trains" of one-second-long presto-chirps (whether ascending and descending), spaced from one another by two second-long silences. Presto-chirps, in each train, are spaced by three semitones, as can be seen in Table 6, which reports the information about the first six notes of this stimulus. The use of a so composed stimulus is only intended to be used to shed a light on how the presto-chirps method deals with long RIRs, and how "grouped" presto-chirps can be used to perform measurements. A discussion of this latter issue is also given in Appendix B.

Pitch	Amplitude	Start Time (sec)	End Time (sec)
21	127	0	1
22	127	3	4
23	127	6	7
24	127	0	1
25	127	3	4
26	127	6	7

Table 6 – Table of parameters used to construct the presto-chirps, extrapolated from the MIDI file of the stimulus.

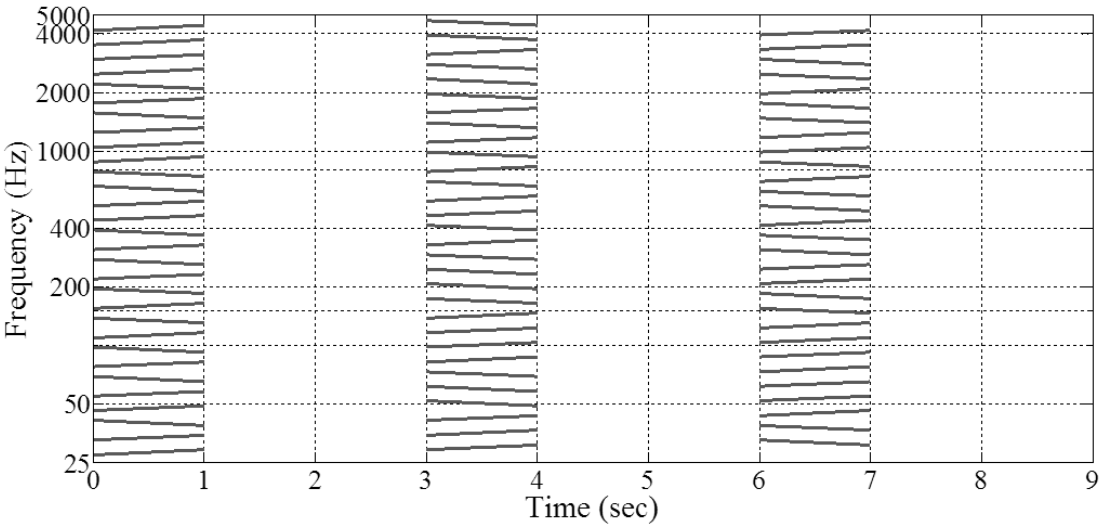


Figure 45 - Stimulus formed by a triplet of 1-second long presto-chirps spaced by two seconds of silence. The presto-chirps of each group are frequency spaced by 3 semitones.

4.2.2.1 Acoustic Parameters Analysis

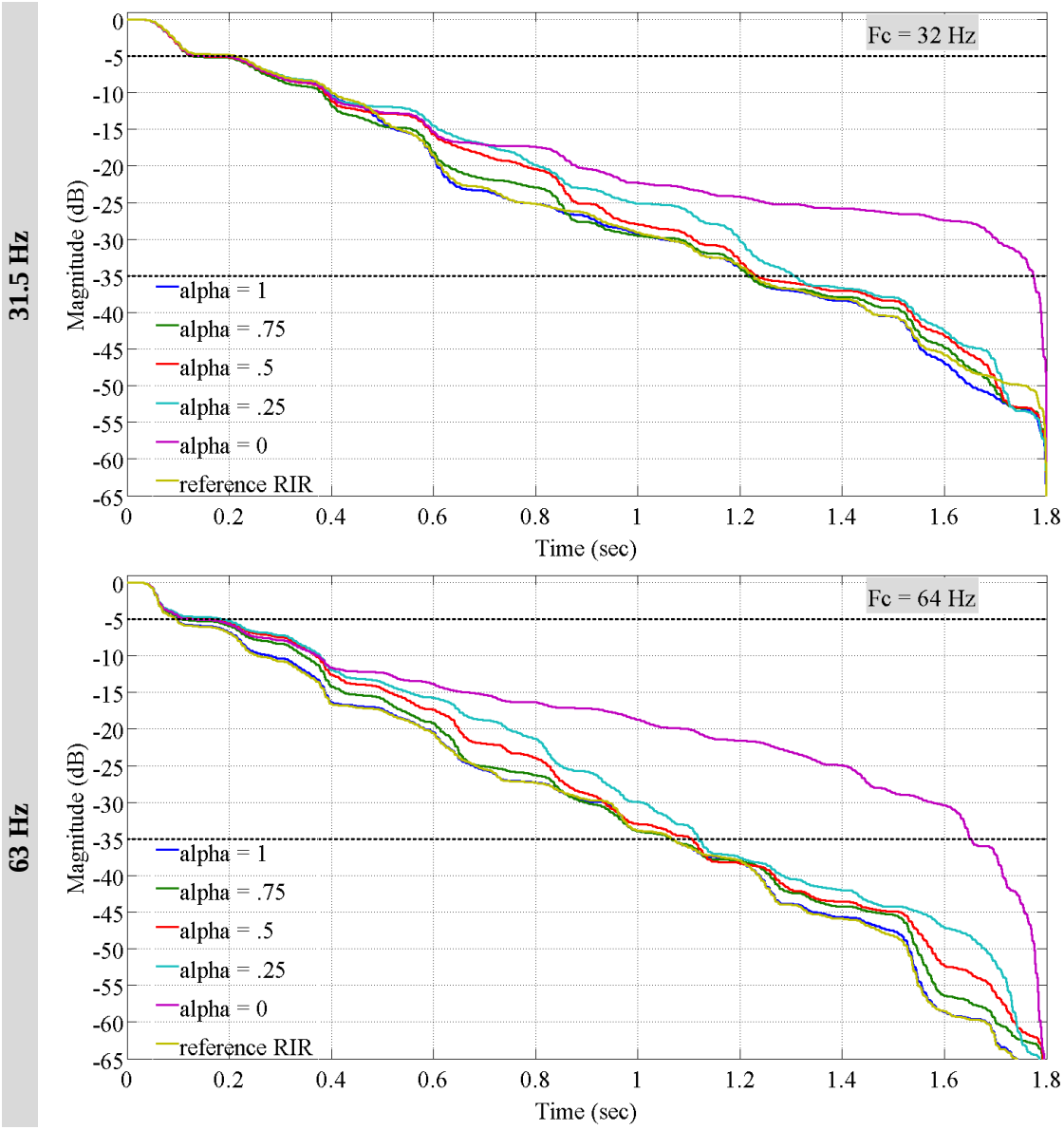
Table 7 shows the reference acoustic parameters, and the differences between those acoustic parameters and the ones calculated from the RIR measured using the presto-chirps method. For $\alpha=1$, all the presented acoustic parameters are within difference limens with the reference values. For $\alpha\neq1$, at the 31.5 Hz and 63 Hz octave bands, all the acoustical parameters are greater than the DLs. Moreover, the parameter D50 is strongly affected also at the 125 Hz and 250 Hz bands. The EDT is inaccurate up to the 250 Hz octave band; the RT30 gives slightly more accurate results, but not for the $\alpha=0$ case. The fact that the acoustic parameters in the lower bands are inaccurate, but for $\alpha=1$, is related to the presence of ripples in the spectra of presto-chirps (as shown in Section 3.1.2.). Ripples contribute to the added (synthesis) noise in the measured RIR, which cause fluctuations of the EDC as shown in the following.

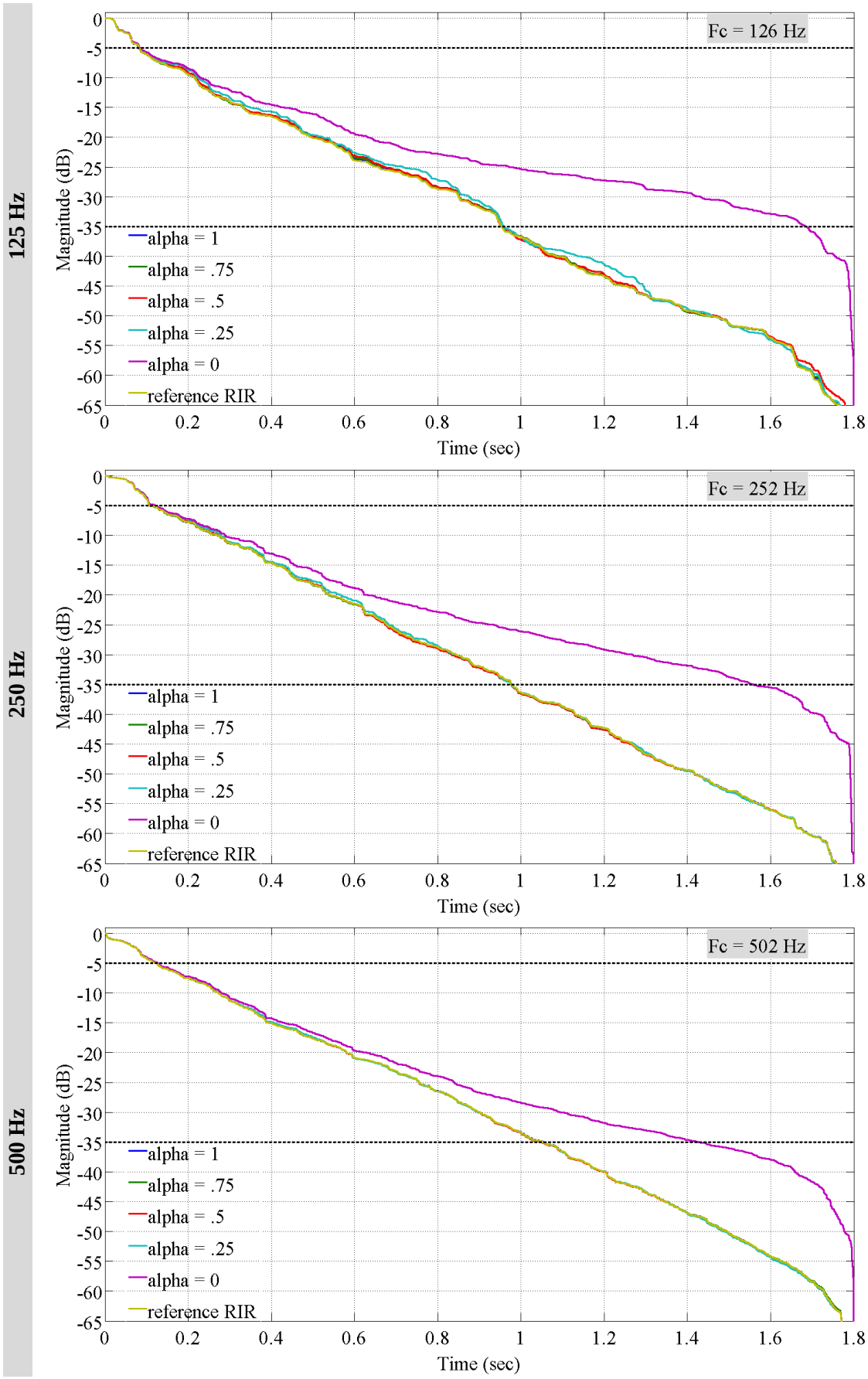
	Bands [Hz]	31.5	63	125	250	500	1k	2k	4k*
C80 [dB] *(DL = 1 dB)*	Reference [sec]	1.07	4.60	4.46	0.25	0.52	1.41	3.67	7.68
	α = 1	0.17	0.12	0.03	0.05	0.04	0.01	0.06	0.05
	α = 0.75	2.17	0.48	0.05	0.16	0.26	0.07	0.06	0.06
	α = 0.5	2.17	1.10	0.43	0.29	0.21	0.01	0.05	0.05
	α = 0.25	2.26	1.66	0.61	0.47	0.04	0.16	0.05	0.04
	α = 0	2.26	1.89	0.69	0.65	0.07	0.36	0.16	0.10
	Bands [Hz]	**31.5**	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
D50 [%] *(DL = 5%)*	Reference [sec]	20.46	69.81	56.37	18.28	28.10	47.62	61.77	79.87
	DL 5% [%]	*1.023*	*3.49*	*2.82*	*0.914*	*1.41*	*2.38*	*3.09*	*3.99*
	α = 1	0.28	0.47	2.97	0.20	0.39	0.10	0.28	0.24
	α = 0.75	43.59	4.45	5.41	3.40	0.07	0.99	0.28	0.24
	α = 0.5	43.56	6.90	3.90	3.30	0.50	0.78	0.20	0.19
	α = 0.25	43.94	9.29	3.25	3.03	1.38	0.27	0.12	0.08
	α = 0	43.77	10.26	3.00	2.50	2.09	1.38	0.81	0.65
	Bands [Hz]	**31.5**	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
EDT [sec] *(DL = 3.2% ± 0.6%)*	Reference [sec]	2.28	1.90	1.26	1.41	1.69	1.70	1.70	1.20
	DL 3% [sec]	*0.06*	*0.05*	*0.03*	*0.04*	*0.04*	*0.04*	*0.04*	*0.03*
	α = 1	0.01	0.03	0.00	0.03	0.00	0.01	0.00	0.00
	α = 0.75	0.20	0.17	0.05	0.01	0.02	0.02	0.01	0.00
	α = 0.5	0.20	0.57	0.17	0.02	0.03	0.03	0.00	0.00
	α = 0.25	0.11	0.88	0.20	0.04	0.04	0.04	0.01	0.01
	α = 0	0.23	1.00	0.22	0.10	0.06	0.04	0.04	0.03
	Bands [Hz]	**31.5**	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
RT30 [sec] *(DL = 3.2% ± 0.6%)*	Reference [sec]	2.12	1.90	1.93	1.72	1.90	1.79	1.65	1.39
	DL 3% [sec]	*0.06*	*0.05*	*0.05*	*0.04*	*0.05*	*0.04*	*0.04*	*0.03*
	α = 1	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	α = 0.75	0.07	0.03	0.04	0.01	0.00	0.00	0.00	0.00
	α = 0.5	0.07	0.09	0.05	0.02	0.02	0.01	0.01	0.00
	α = 0.25	0.18	0.17	0.09	0.02	0.04	0.02	0.01	0.01
	α = 0	2.23	2.18	0.16	0.14	0.14	0.12	0.12	0.01

Table 7 – Acoustic parameters estimation in the octave bands from 31.5 Hz to 2 kHz, and in the third octave band centred at 4 kHz - for several α.

Figure 46 reports the EDCs in the octave bands 31.5 Hz – 4 kHz. In the 31.5 Hz and 63 Hz bands the only α=1 curve matches the reference EDC, with the curve α=1 being strongly bended by the synthesis noise, in all the sub bands. For 1<α<0 the EDCs show slight fluctuations up the 500 Hz band, and are generally equal to the reference

EDC in the higher bands. This indicates the effect of the synthesis noise is larger at lower bands due to the still high magnitude ripples in the spectra of presto-chirps.




125 Hz
250 Hz
500 Hz
Fc = 126 Hz
Fc = 252 Hz
Fc = 502 Hz
Magnitude (dB)
Time (sec)
alpha = 1
alpha = .75
alpha = .5
alpha = .25
alpha = 0
reference RIR

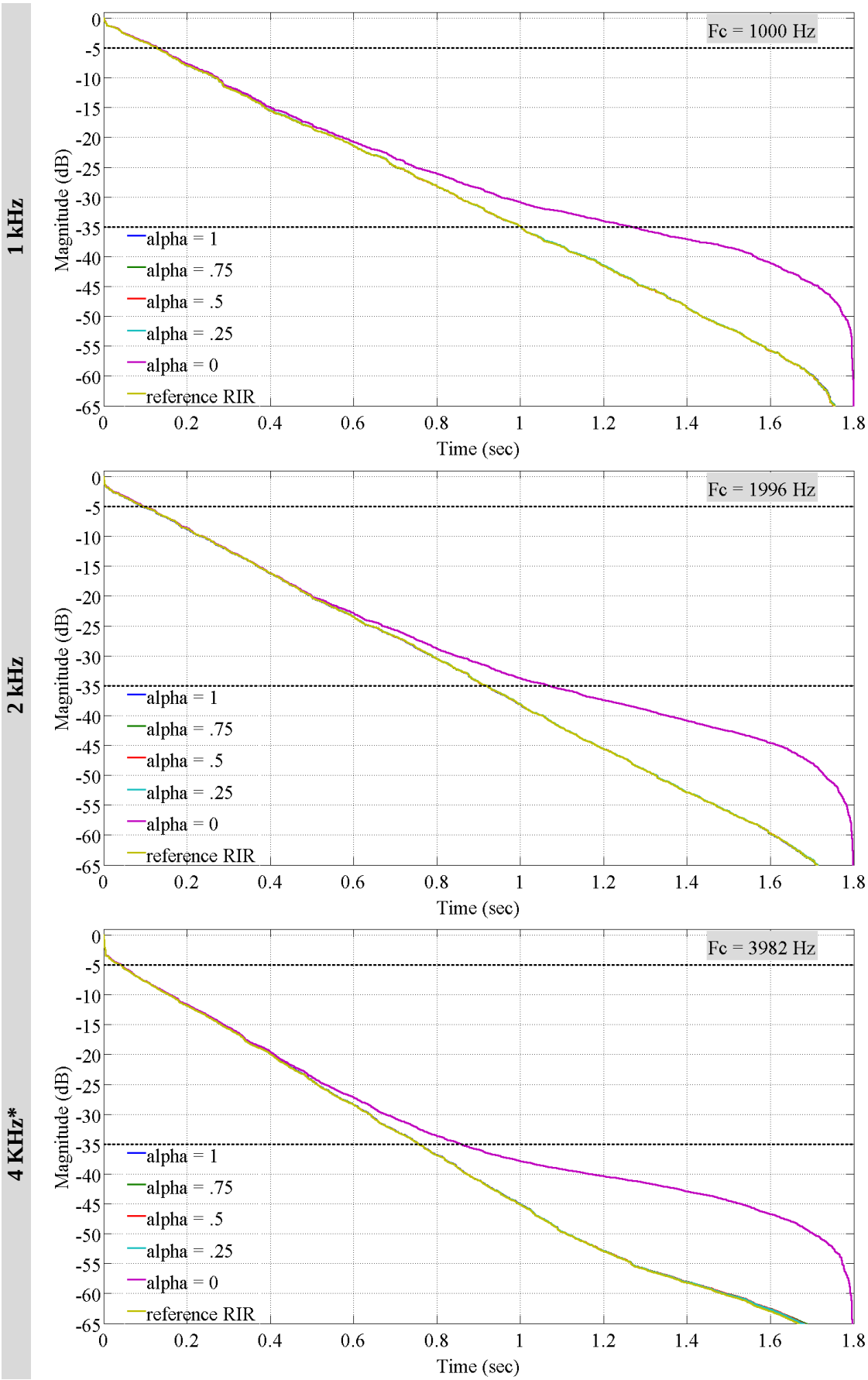
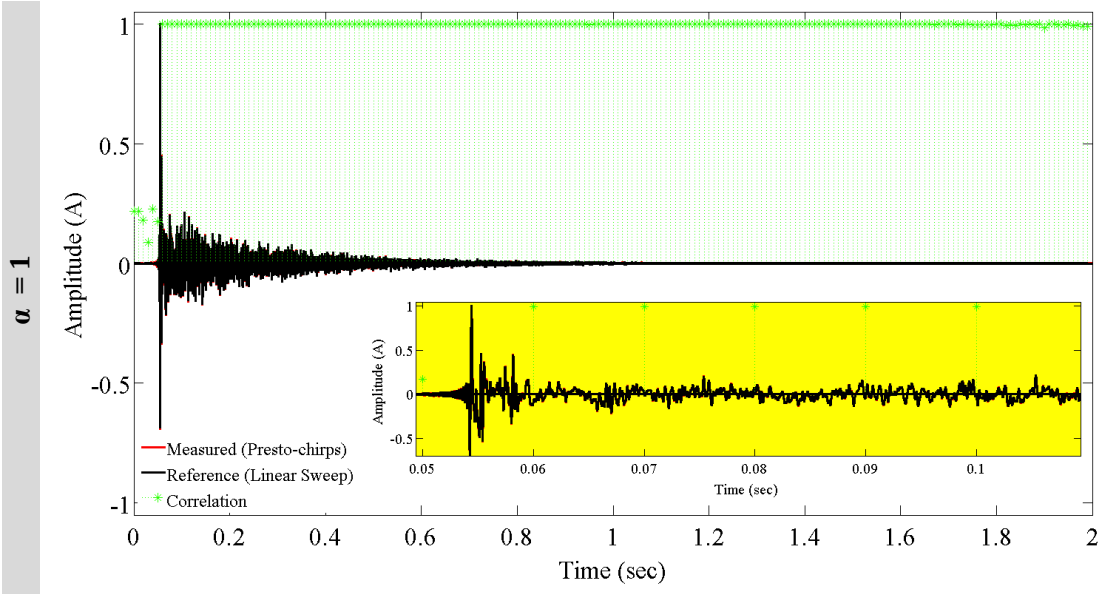


Figure 46 – EDCs of the measured RIRs (as function of α) plus the reference RIR. Each figure accounts for a different sub band as labelled in the left. The dashed lines indicate

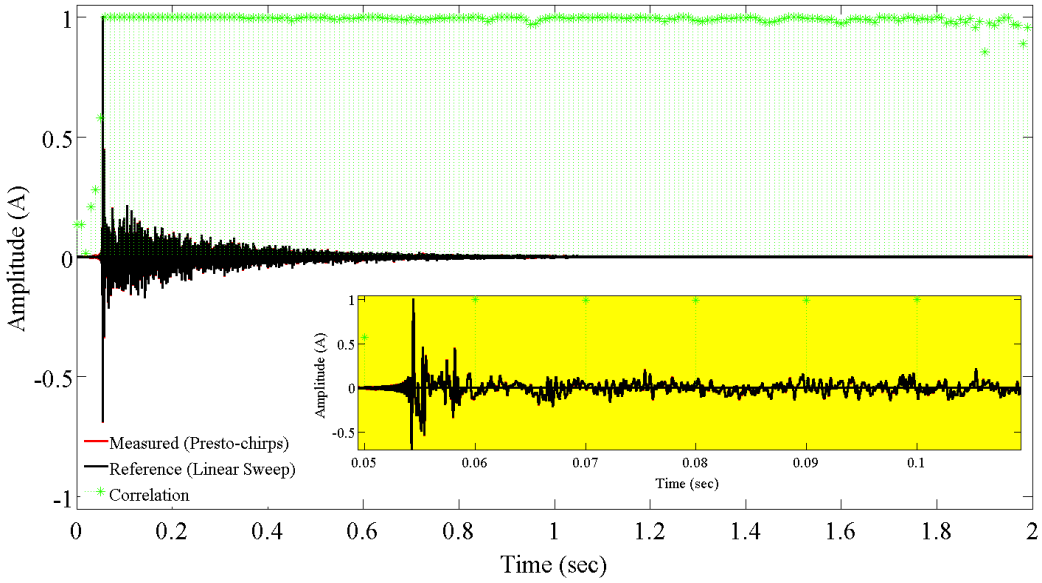
the decay range, which is commonly used to calculate the acoustic parameter RT30. The label 'Fc' represents the central frequency of the octave band filter used.

4.2.2.2 Time Domain Analysis

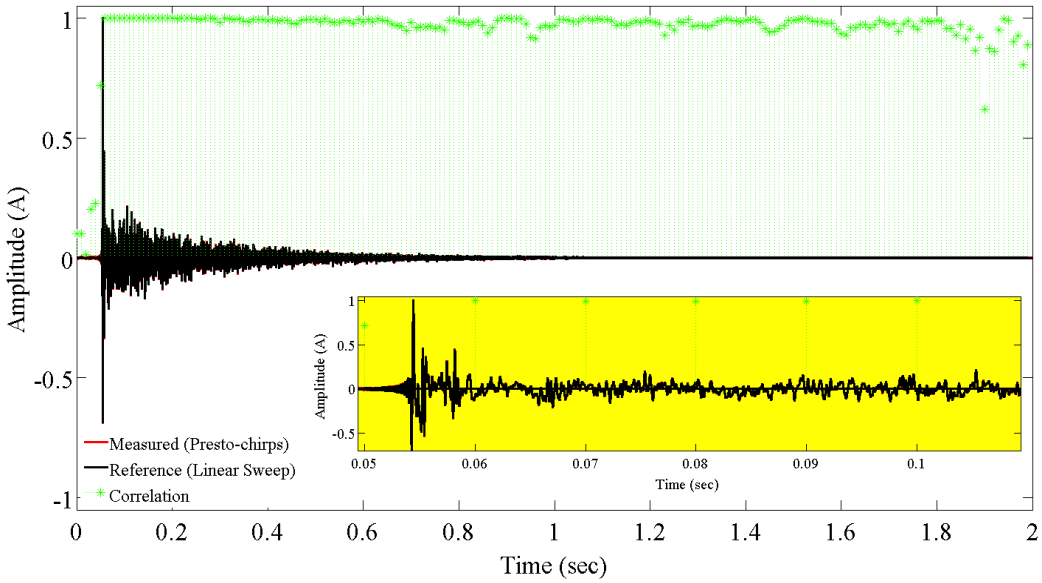
Figure 47 highlights that for α=1 the measured RIR correlates almost perfectly with the reference RIR. For α < 1, the correlation becomes progressively worse, starting from the tail of the RIR, and then progressively towards the main peak as α decreases. Note the presence, before the main peak, of the synthesis noise in the measured RIR and how it increases for lower α values.



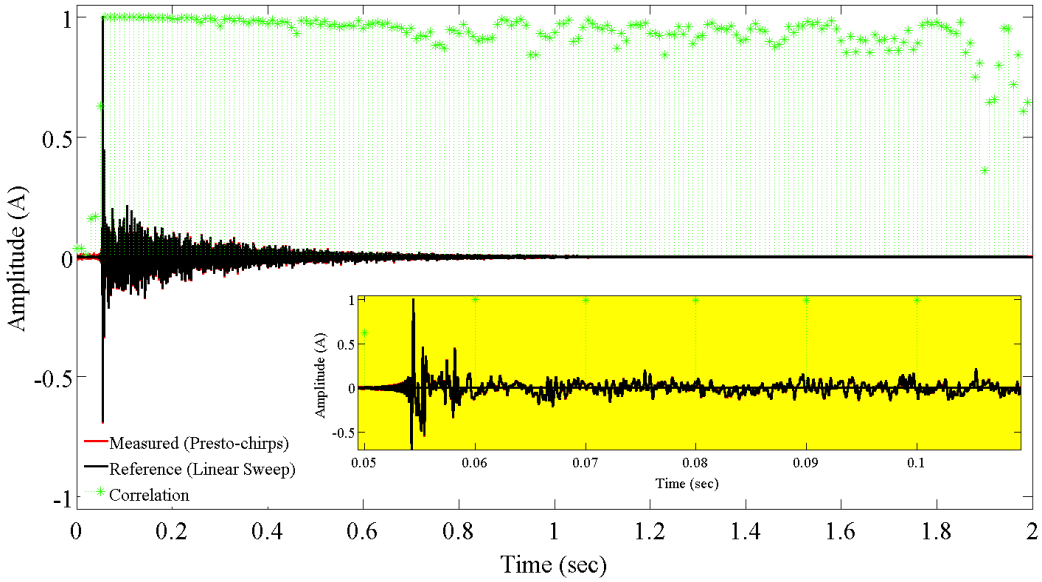
α = 0.75



α = 0.5



α = 0.25



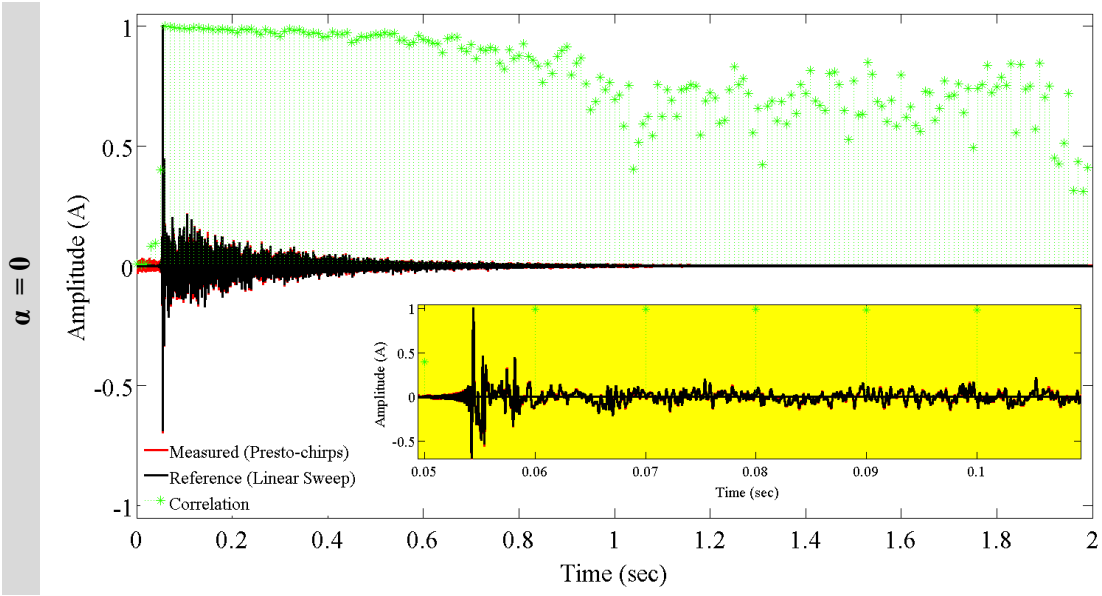
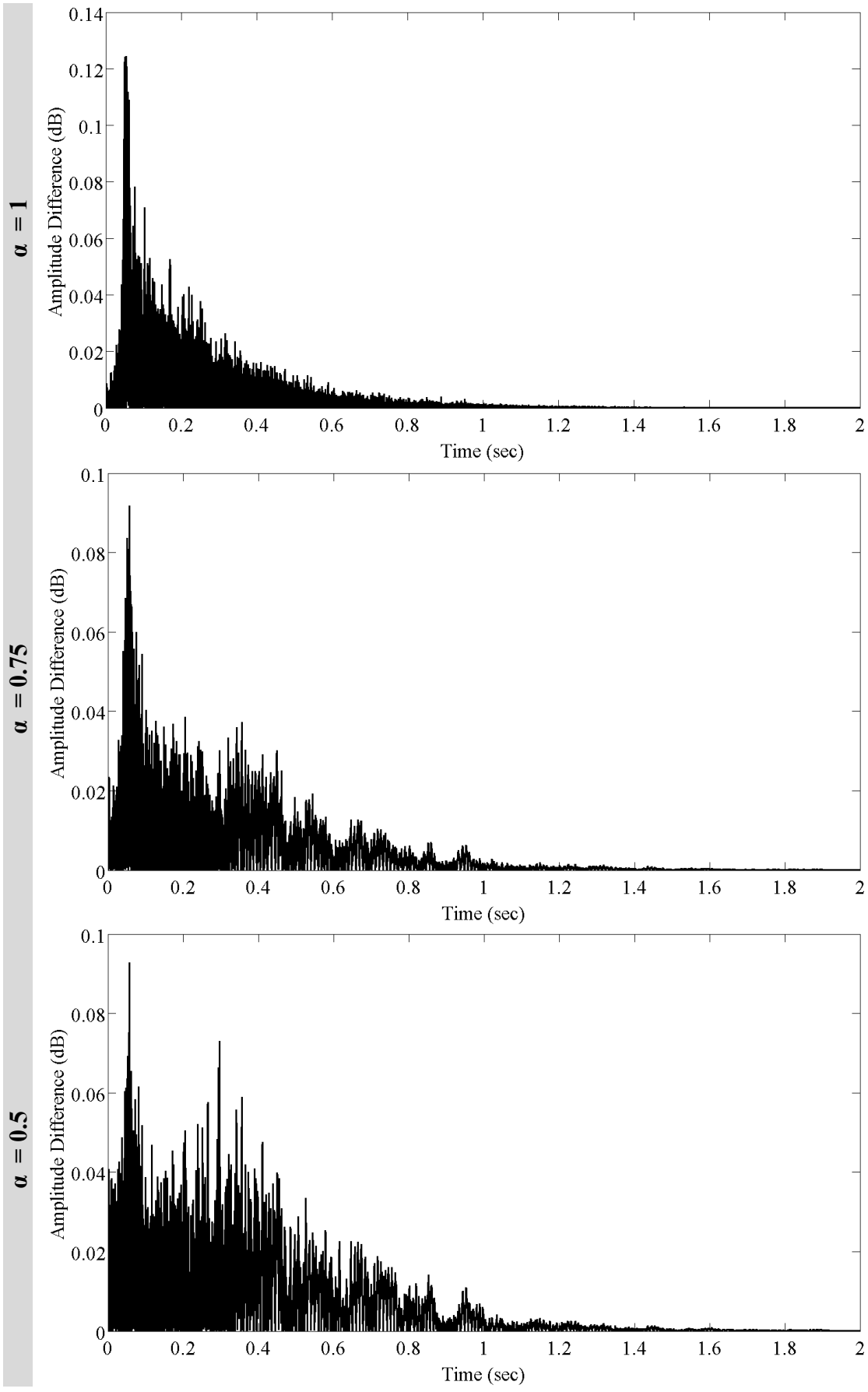


Figure 47 – Magnitude differences functions for several α values.

4.2.2.3 Amplitude Differences

Figure 48 shows the amplitude differences between the reference RIR and the RIRs measured in the simulations. Such bias is only due (since the tested RIR is time invariant and noise-free) to the ACF, where the side lobes interact with the side lobes of neighbouring peaks. Differences are lower for the α=1 case, due to the fact that in such a case the ACF of the stimulus presents short and low-level side-lobes (as shown in Figure 34). In contrast, for $\alpha \neq 1$ the higher amplitude differences are spread along the RIR, instead that around the main peak.


α = 1
Amplitude Difference (dB)
Time (sec)
α = 0.75
Amplitude Difference (dB)
Time (sec)
α = 0.5
Amplitude Difference (dB)
Time (sec)

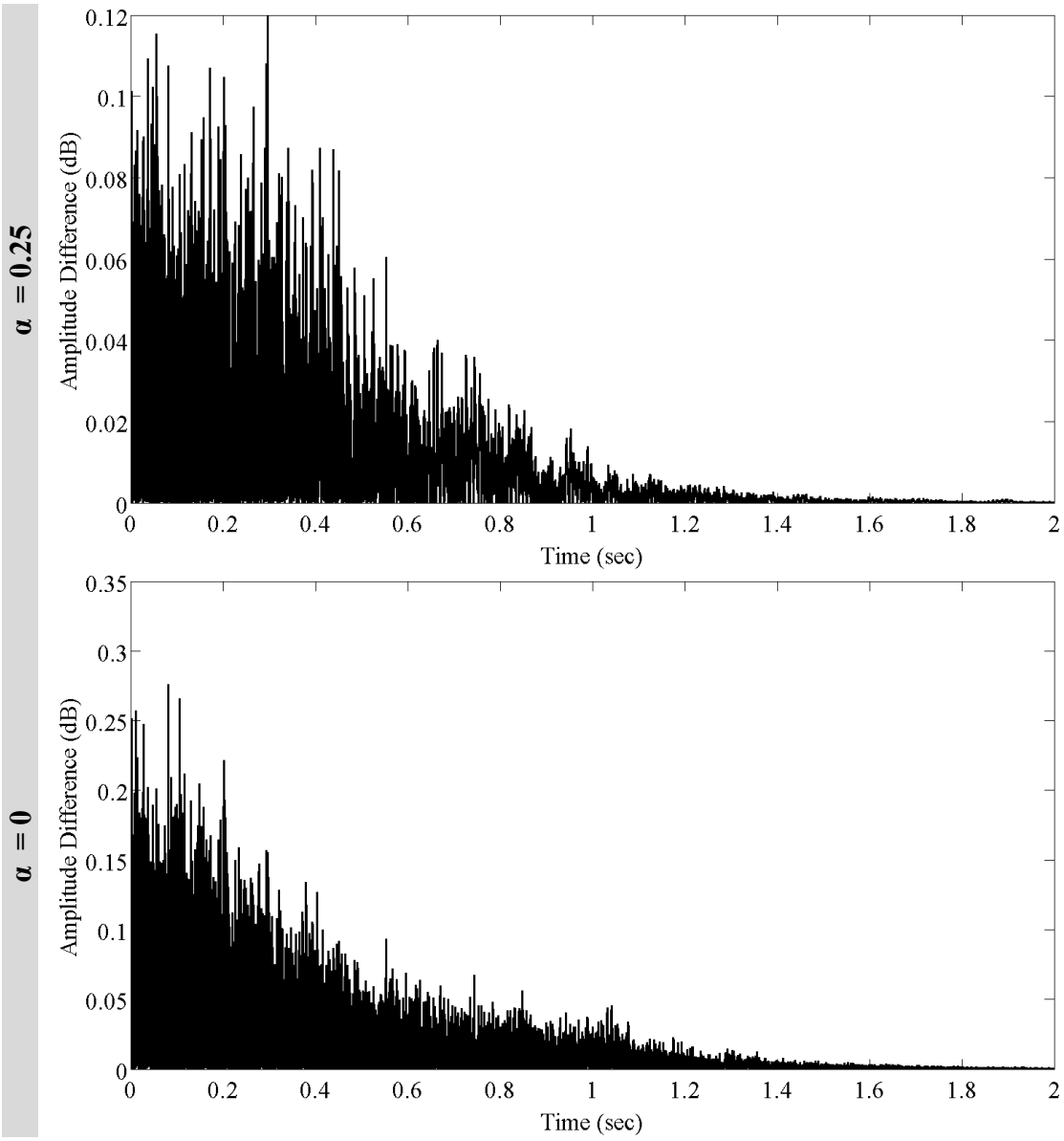
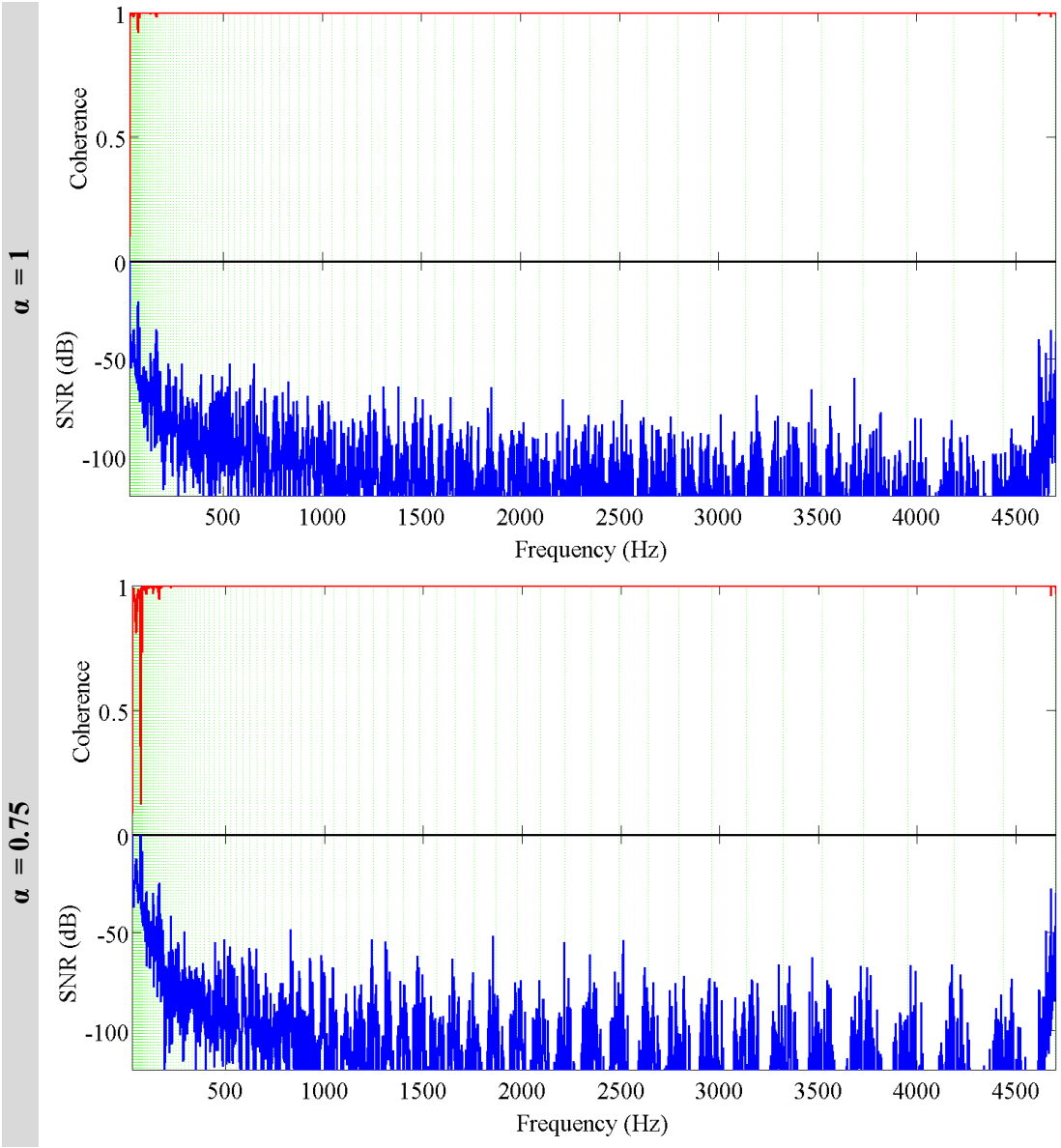


Figure 48 – Amplitude differences versus α.

4.2.2.4 Frequency Domain Analysis

Figure 49 shows that for α=1 the coherence is equal to one for the whole measured spectrum, except for few frequencies at the low end of the spectrum. For lower α values the coherence decreases, starting from the low frequency bands and then spreading up to the high end of the spectrum. The artefacts in both the coherence and SNR functions, are localised around the "junction points" between presto-chirps

(green vertical lines). Interactions between the side lobes of presto-chirps give the specific noise distribution shown in the figures.



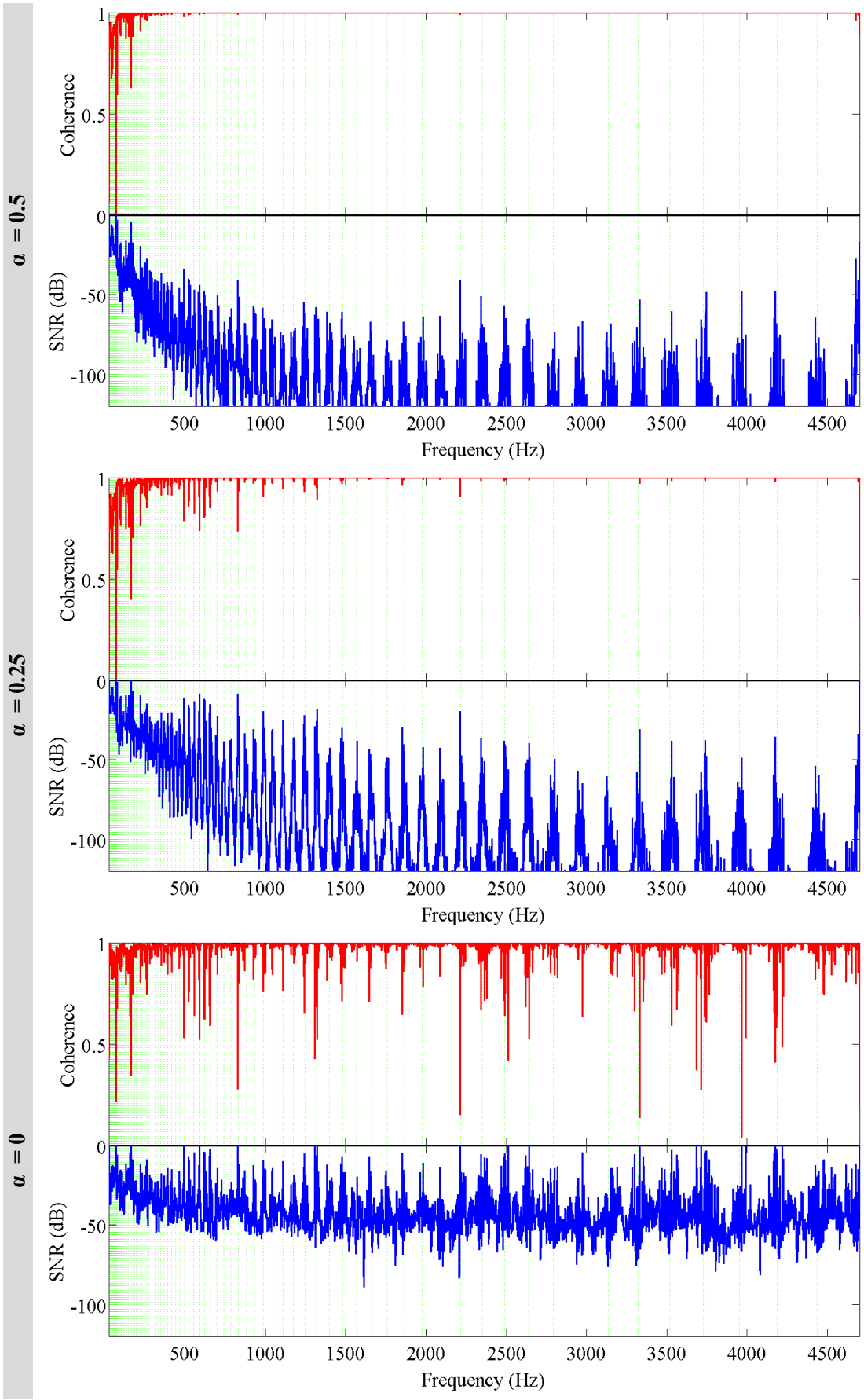
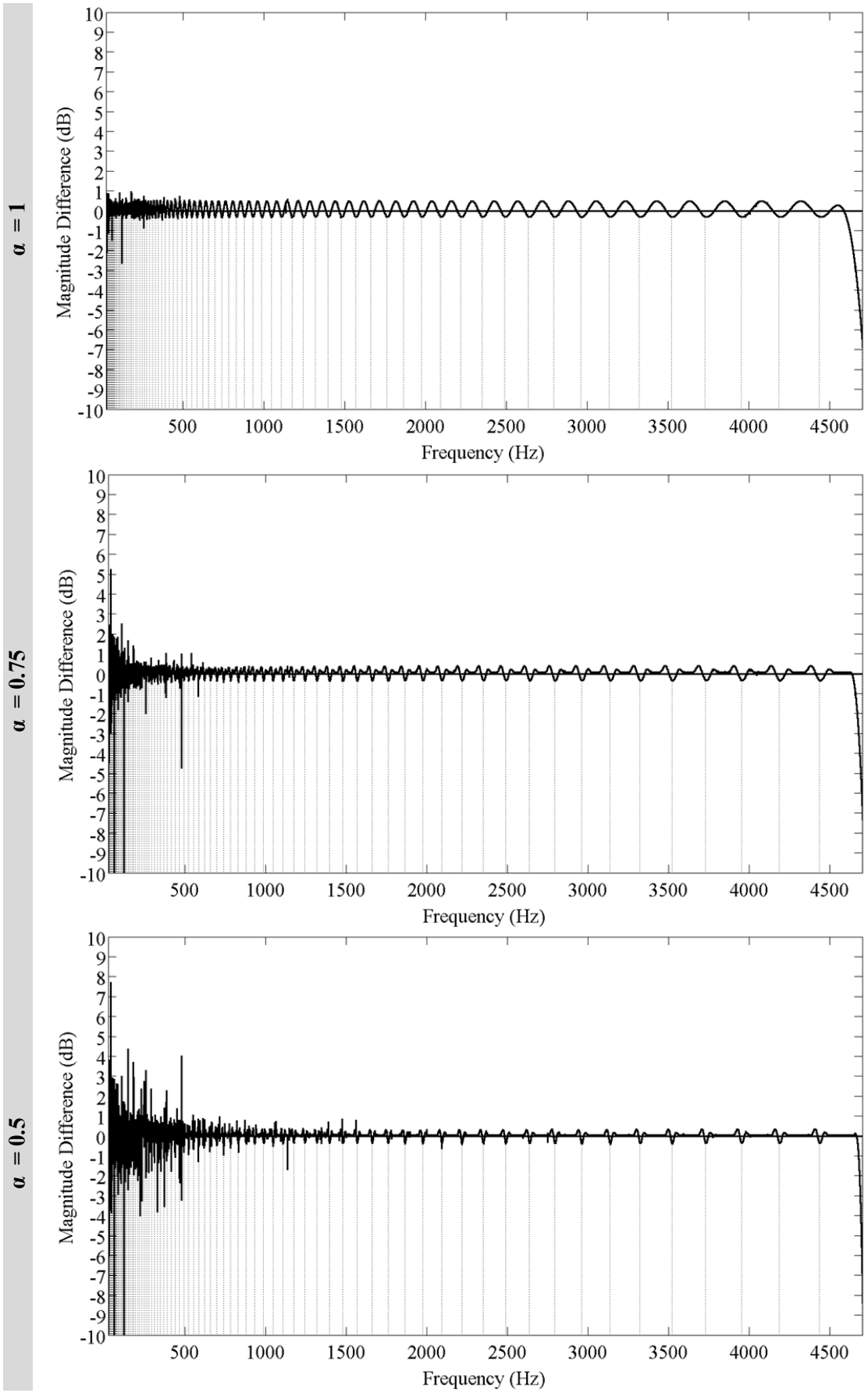


Figure 49 – Coherences function as a function of α.

4.2.2.5 Magnitude Differences

Figure 50 shows the magnitude differences between the reference RIR and the ones measured using the presto-chirps method for the labelled α values. These comparisons shed a light on how the frequency ripples of the presto-chirps (described in Section 3.1.2) produce increasingly higher differences as α lowers. The effect is more marked if the measured FRF has local minima near the junction points. This has been deducted by noting the presence of spike-like peaks at some frequency points along the FRFs. As a result, the distribution and magnitude of the errors depend on both the analysed FRF, and on the used window. For example, the measurement of an FRF with a maximally flat frequency response (e.g. like the one of a low pass filter discuses in Section 4.2.1), will result in a magnitude difference function equal to the bias introduced by the stimulus (since there are not minima in the FRF). Hanning-windowed presto-chirps present additional magnitude errors at the low frequency bands, where the pass bands of the presto-chirps interact considerably. Therefore, the measured FRF is closer to the reference one than in the other cases. Note that since in these presented measurements where one-second long presto-chirps were used, the averaging effect discussed in 4.2.1.6 is not present.


α = 1
α = 0.75
α = 0.5
Magnitude Difference (dB)
Frequency (Hz)
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
500
1000
1500
2000
2500
3000
3500
4000
4500

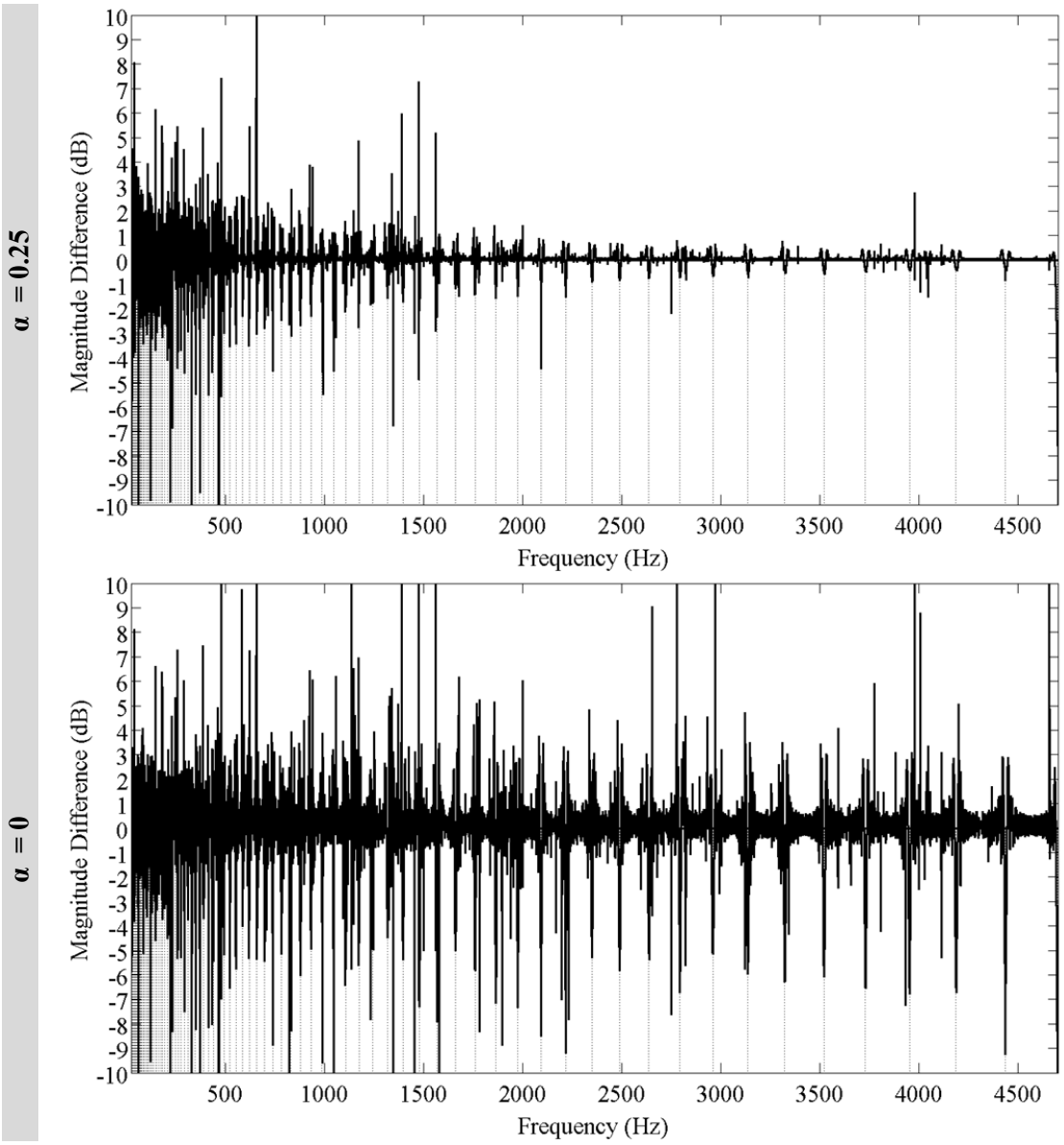


Figure 50 – Magnitude differences functions for several α values.

4.2.2.6 Discussions on the Case Study

The acoustic parameter D50 shows greater discrepancies from the reference values, compared with the other parameters. Other values outside the DLs can be noted at low frequency bands, for low α, for the acoustic parameter RT30. The correlation between the whole reference and the measured RIRs, reported in Figure 51, showed that for α=1 the RIRs is virtually one (total correlation). The other window configurations

instead introduce, as α decreases, progressively more synthesis noise, which decreases the correlation.

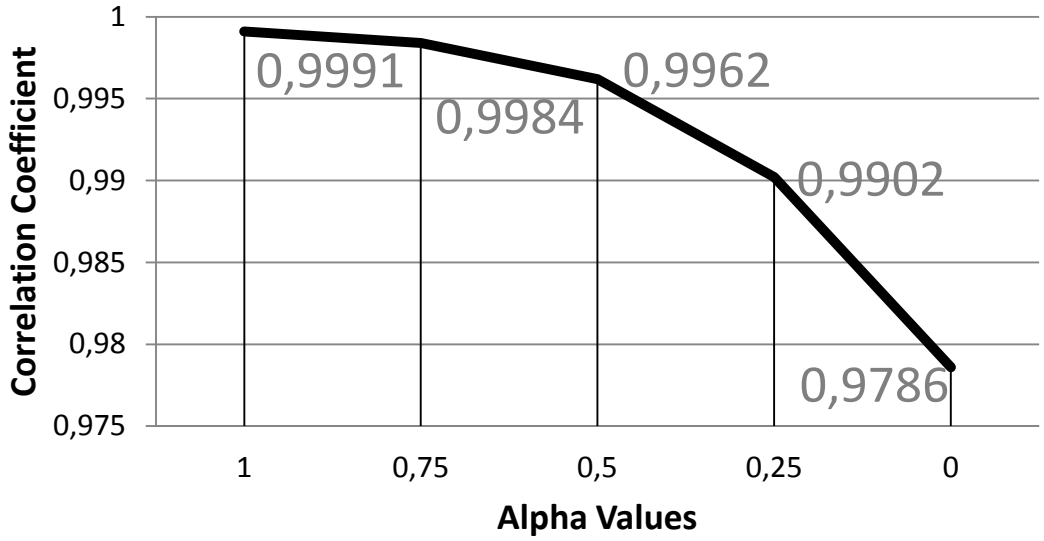


Figure 51 – Degree of correlation between the reference RIR and the RIRs measured with the different α-windowed stimuli.

The synthesis noise and the frequency domain bias are strictly related, being one the Fourier transform of the other. Narrower bandwidths have higher side lobes, which translate in ACFs with prolonged ringing. This increases the likelihood that the side lobes of neighbouring peaks interact with one another, producing as a result, a diffuse and high-level noise along the RIR. This explains the bending of the EDCs for α=0, as shown in Figure 46. Coherence analyses have shown that the introduced bias is minimal for α=1. It progressively increases as α decreases. The frequency bias is higher at the lower frequency bands, where the narrow bandwidths of presto-chirps cause them to have higher side-lobes. Side lobes interact with one another and this causes greater magnitude differences for lower α. Moreover, at the junction points between presto-chirps, high side lobes leak over the local minima of the measured FRF, causing estimation errors.

4.3 In Situ Measurements

In situ measurements have been carried out in three different environments, to prove the accuracy of the presto-chirps method to deal with real systems. Impulse response measurements were performed using the stimulus presented in Section 3.6, for the Tukey window configurations α = 1, 0.75, 0.5, 0.25, and 0. This section only reports one of three tested venues. The other two can be found in Appendix C. The measurement setting was equal in all the tested scenarios. It is described in the following.

4.3.1 Measurements Setting

The reference impulse responses have been measured using a 10 seconds long linear sine sweep ranging from 10 Hz to 5 kHz, and then filtered in the range 27 – 4.7 kHz, to match the frequency range of the presto-chirps method. Both the reference RIR and the presto-chirps stimuli were played with a level of 70 dB (L_{Aeq}), calculated over the length of the signal. The measurement chain was composed of a personal computer "Arqivia" (2.8 GHz Pentium 6), equipped with the external soundboard "M-audio – fast-track 400". The technical details of the used loudspeaker and microphone are reported below:

Loudspeaker: Genelec 1030A:

- Frequency range (free field), ± 2.5 dB: 55 Hz - 18 kHz:
    - Lower cut-off frequency, –3dB: < 52 Hz
    - Upper cut-off frequency, –3dB: > 20 kHz
- Maximum long term RMS acoustic output @ 1m > 99 dB SPL
- Maximum short term sine wave acoustic output @ 1m > 105 dB SPL

Microphone: Bruel&Kaier DPA 4007 (omnidirectional):

- Frequency range, ± 2 dB: 10 Hz - 40 kHz
- Equivalent noise level, A-weighted: Typ. 24 dB(A) re. 20 µPa (max. 26 dB(A))
- Total harmonic distortion (THD): < 0.5 % up to 142 dB SPL peak
- Dynamic range: Typ. 124 dB Max. SPL, peak before clipping: 160 dB

As mentioned in Section 4.1, the point-point magnitude and the amplitude difference functions, used in the simulated scenarios, are not used in these validations. The mixed-phase and time-variant nature of RIRs will not give equal responses, even for measurements carried out in consecutive periods. To prove it, the early part of two RIRs, measured one after another using the standard ESS technique, produced the impulse and the frequency responses depicted in Figure 52 and Figure 53, respectively. In both cases, slight differences in the responses are clearly visible.

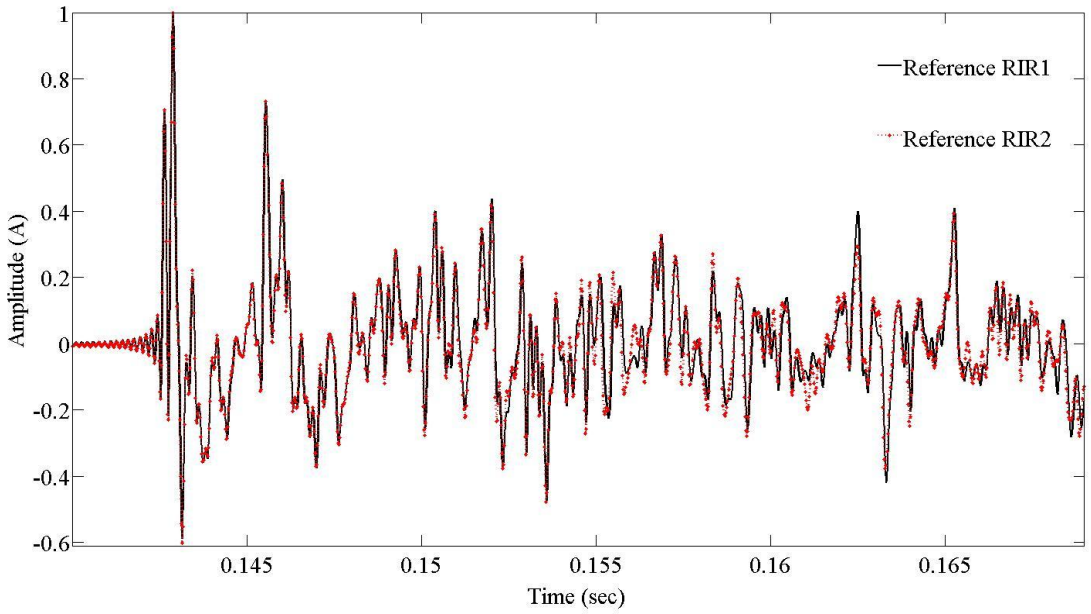


Figure 52 – Comparison of two RIRs measured using the ESS method.

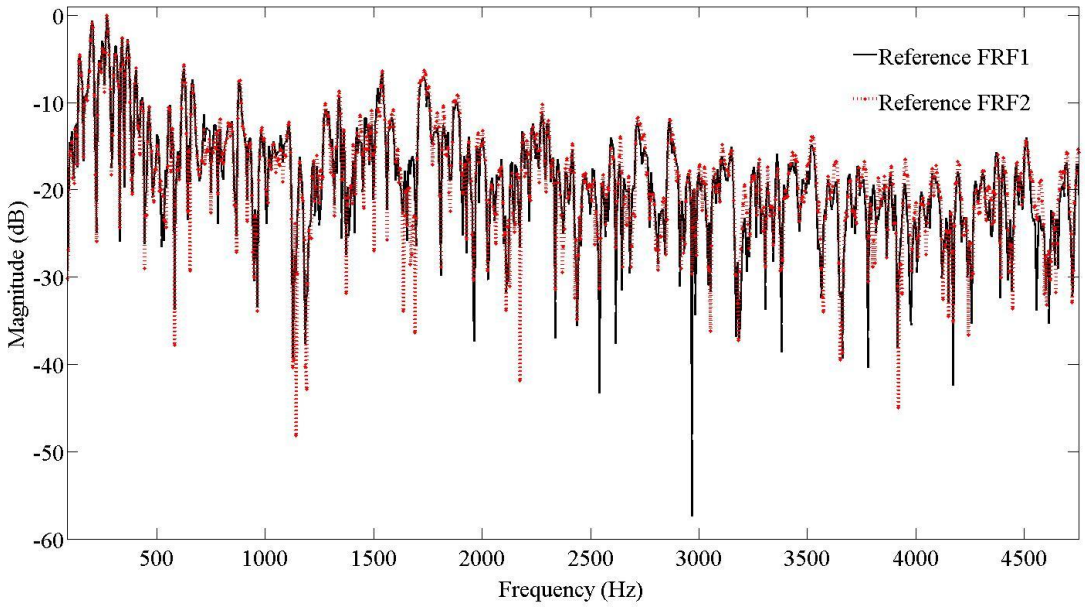


Figure 53 –FRFs of the RIRs shown in Figure 52.

4.3.2 Case Study: Broadcasting Auditorium "Sala B"

This case study concerns the impulse response measurements of a medium-sized rectangular-shaped auditorium. This auditorium is used for live broadcasting radio programs, which include the presence of an audience (around 100 people, included the technical staff and the musicians). Present In the room is a stage (around 100 square meters), which is used by live bands and hosted artists. In front of the stage there is the table used by the radio presenters.

Figure 54 shows a photo of the auditorium, taken from the stage, while Figure 55 shows its architectural drawing, and reports the positions of the microphone and the loudspeaker. Figure 56 shows, in the upper panel, the measured (reference) RIR of the auditorium, while its frequency response in reported in the lower panel. The auditorium reverberation time RT30, averaged within the sub bands from 63 Hz to 4 kHz, was about 0.86 seconds.

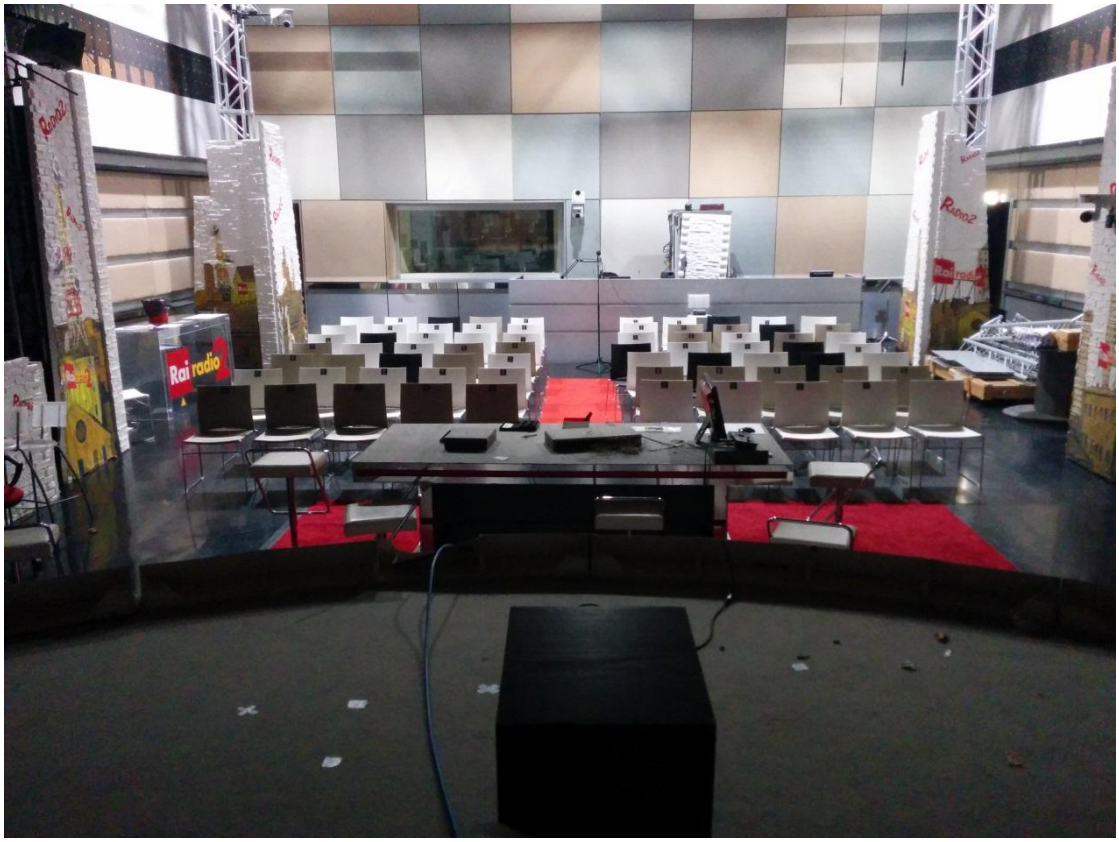


Figure 54 – Photograph of the auditorium “Sala B”.

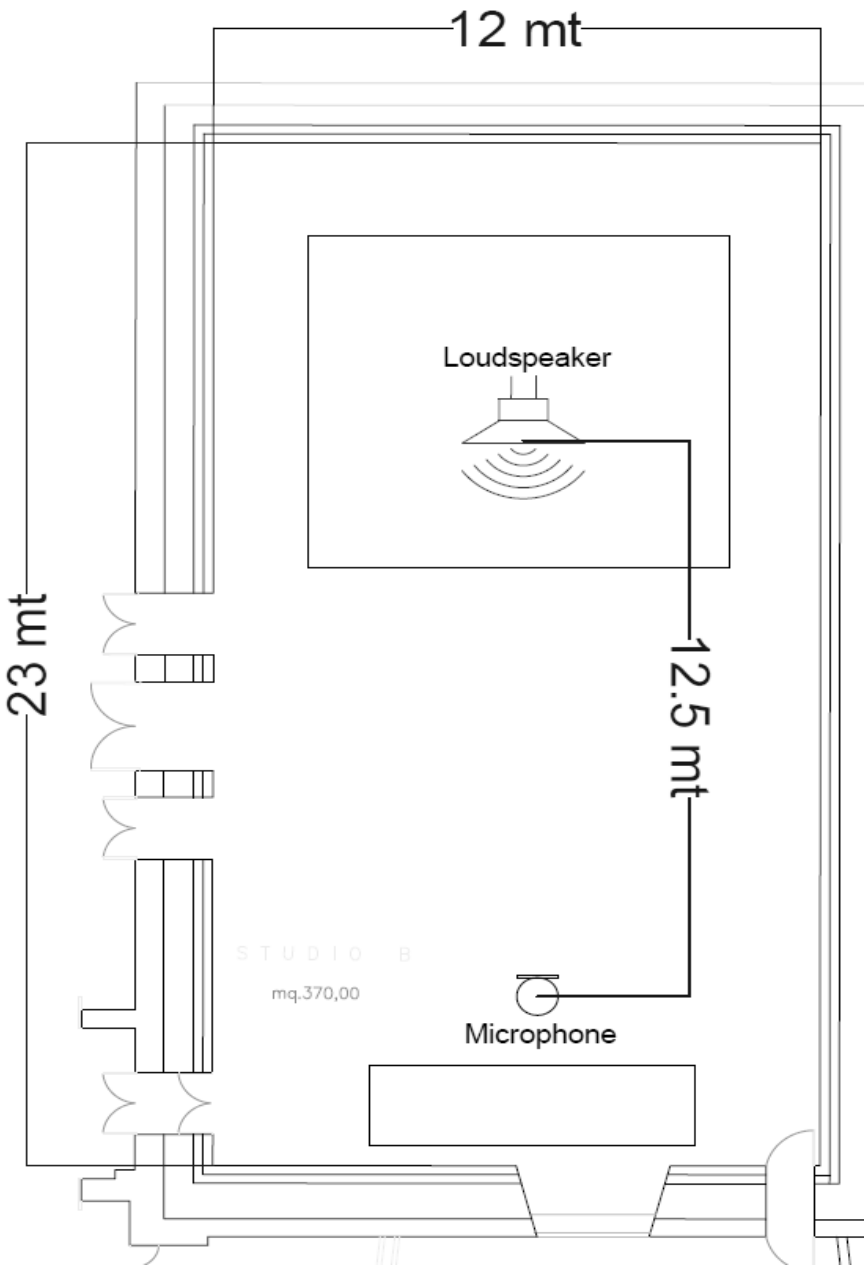


Figure 55 – Architectural drawing of the auditorium "Sala B", with quotas and positions of loudspeaker and microphone.

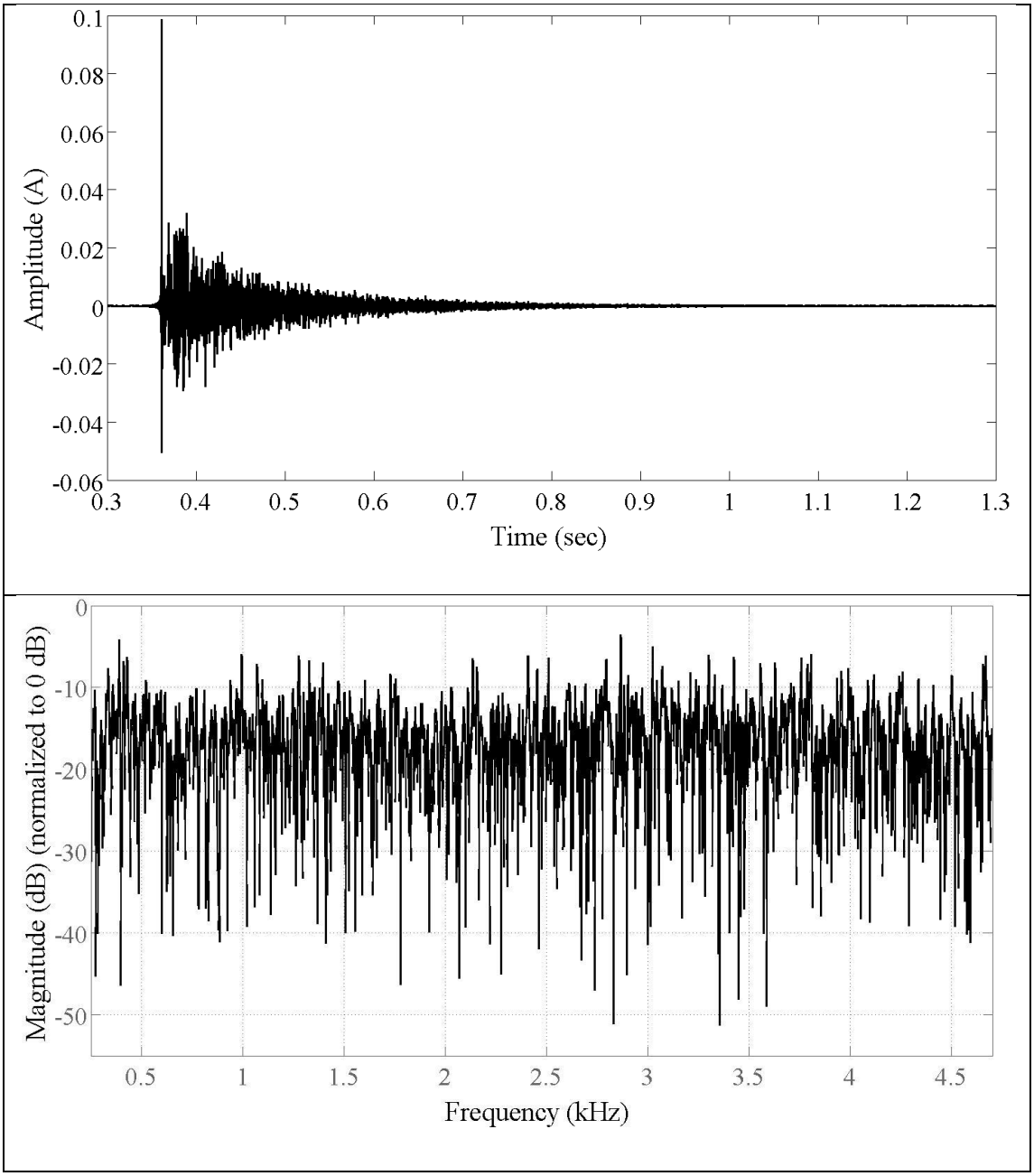


Figure 56 – In the upper and bottom panels are shown the impulse and frequency responses of the tested room, respectively.

4.3.2.1 Acoustic Parameters Analysis

Table 8 reports the comparisons between of the acoustic parameters of the reference measurements and the ones calculated from the RIR measured using the presto-chirps method, for several tested window configurations. Acoustic parameters are calculated using Audacity, with plug-in Aurora. The 31.5 Hz octave band is not reported as the reference measurement had a too low SNR to allow meaningful comparisons. This effect can be seen in its EDC reported in Figure 57. The 63 Hz octave band also shows doubtful results due to the low SNR of the reference measurements.

The acoustic parameter C80 and D50 are within the DLs for all the tested configurations, in the sub bands from and above 125 Hz. The EDT is within DL in the bands from and above 250 Hz, and in the 125 Hz band for α=1. This result is confirmed for all the tested configurations. The RT30 is within the DLs in the sub bands from and above 125 Hz only for α=1. For the other cases, the measured values are within the DLs only from and above the sub bands 1, 2, and 4 kHz.

C80 [dB] *(DL = 1 dB)*

Bands [Hz]	63	125	250	500	1k	2k	4k*
Reference	9.35	5.073	2.536	6.091	7.465	6.401	5.931
α = 1	2.09	0.835	0.23	0.035	0.056	0.092	0.032
α = 0.75	2.71	0.932	0.279	0.08	0.014	0.107	0.032
α = 0.5	3.62	0.661	0.285	0.012	0.237	0.207	0.09
α = 0.25	4.10	0.386	0.175	0.195	0.344	0.121	0.096
α = 0	3.82	0.171	0.116	0.423	0.83	0.636	0.274

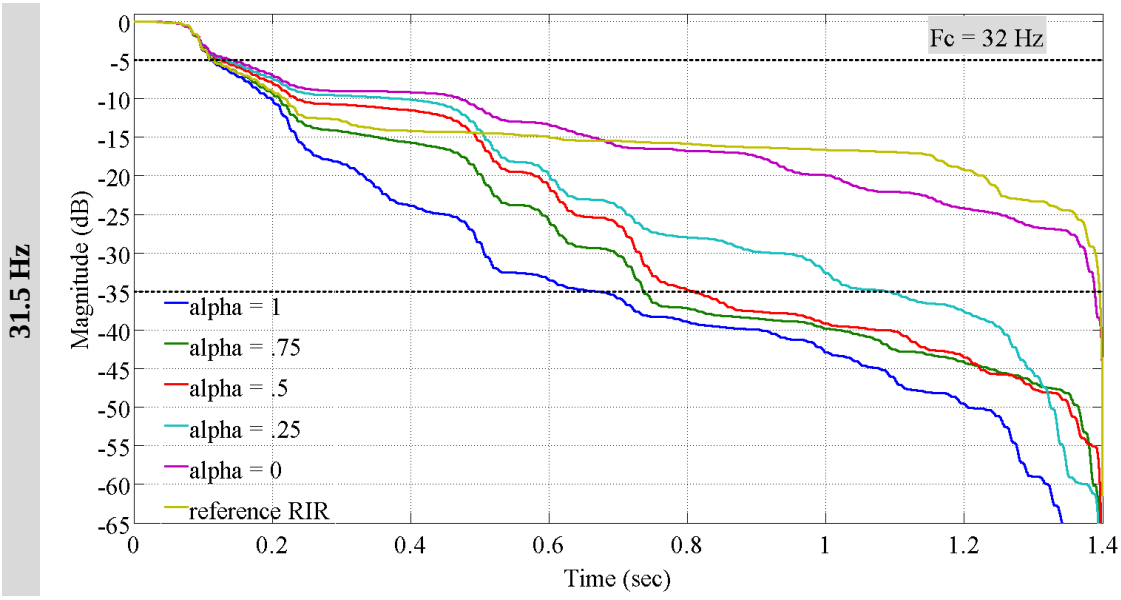
	Bands [Hz]	63	125	250	500	1k	2k	4k*
D50 [%] *(DL = 5%)*	Reference	81.2	66.5	40.44	63.5	68.66	70.25	69.2
	DL 5% [sec]	4.06	3.33	2.02	3.18	3.43	3.52	3.47
	α = 1	4.29	3.02	1.976	0.55	0.949	0.224	0.08
	α = 0.75	5.99	3.12	1.894	0.43	0.611	0.438	0.21
	α = 0.5	8.89	2.10	1.988	1.45	0.011	0.795	0.19
	α = 0.25	10.7	1.12	1.951	2.02	0.162	1.002	0.42
	α = 0	9.45	6.01	1.443	2.36	1.687	1.995	1.04
	Bands [Hz]	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
EDT [sec] *(DL = 3.2% ± 0.6%)*	Reference	0.907	0.862	0.802	0.682	0.682	0.882	0.831
	DL 3% [sec]	*0.015*	*0.026*	*0.024*	*0.020*	*0.020*	*0.026*	*0.025*
	α = 1	0.152	0.020	0.018	0.01	0.006	0.014	0.007
	α = 0.75	0.216	0.146	0.007	0.011	0.001	0.018	0.011
	α = 0.5	0.385	0.117	0.013	0.009	0.009	0.022	0.004
	α = 0.25	0.694	0.085	0.015	0.001	0.012	0.006	0.018
	α = 0	0.515	0.116	0.017	0.013	0.013	0.003	0.025
	Bands [Hz]	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
RT30 [sec] *(DL = 3.2% ± 0.6%)*	Reference	0.974	0.904	0.881	0.833	0.814	0.878	0.875
	DL 3% [sec]	*0.030*	*0.027*	*0.026*	*0.025*	*0.024*	*0.026*	*0.026*
	α = 1	0.188	0.022	0.025	0.025	0.006	0.002	0.004
	α = 0.75	0.473	0.034	0.057	0.044	0.001	0.002	0.004
	α = 0.5	0.637	0.041	0.082	0.056	0.005	0.001	0.001
	α = 0.25	0.767	0.061	0.086	0.079	0.007	0.008	0
	α = 0	0.636	0.126	0.096	0.047	0.006	0.02	0.002

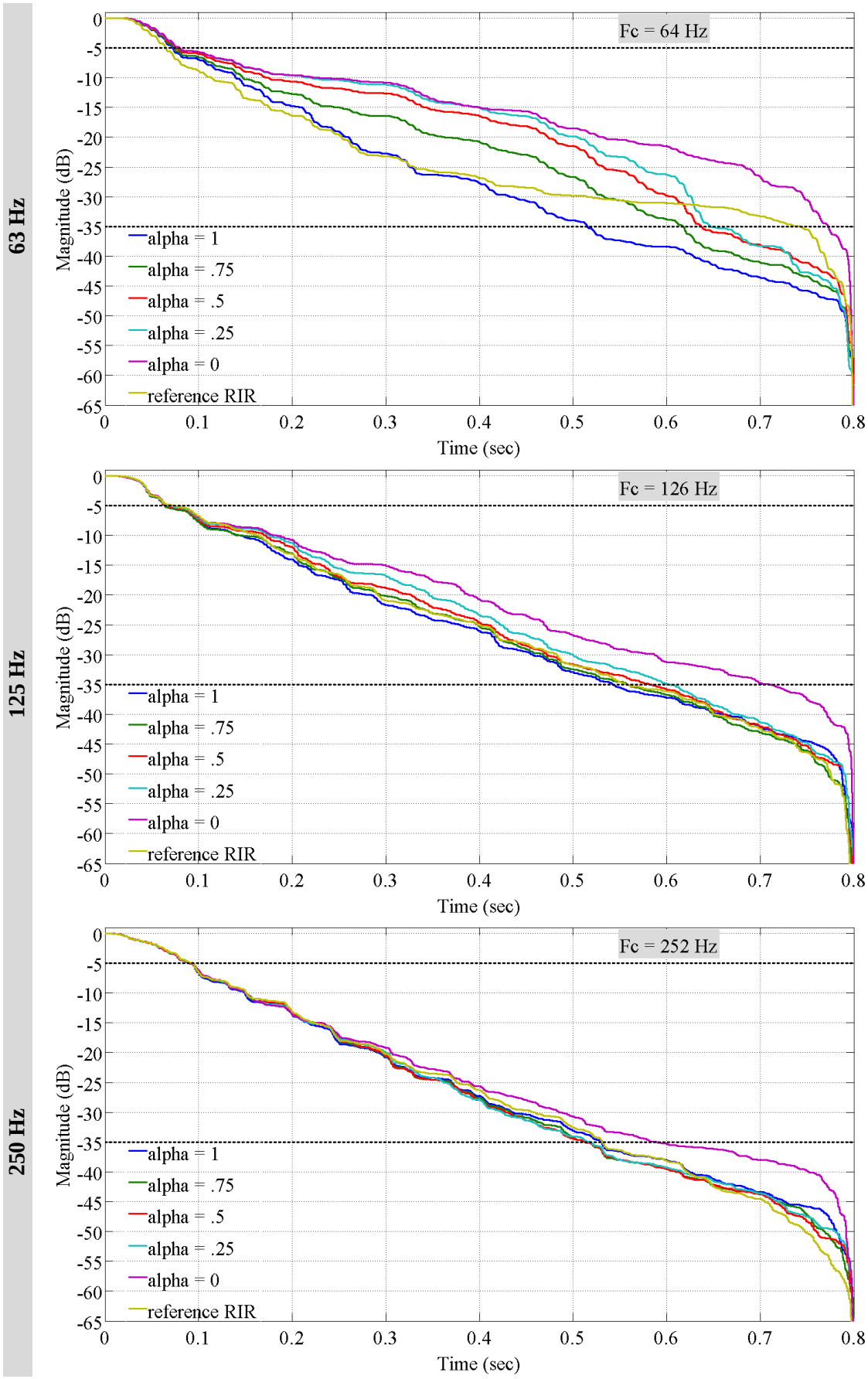
Table 8 – Acoustic parameters estimation in the octave bands from 63 Hz to 2 kHz, and in the third octave band centred at 4 kHz - for several α.

Figure 57 reports the EDCs calculated for the octave bands from 31.5 Hz to 4 kHz; each panel describes an octave band as labelled in the left. the octave band filter central frequencies (Fc) are indicated in the grey box. The dashed horizontal lines, at -5 and -35 dB, indicate the decay range commonly used to calculate the RT30. The EDCs are plotted using different upper integration limits, depending if a sub band was, or was not, affected by deconvolution residuals. In fact, since residuals have specific frequency content (as discussed in Appendix A), they will not "appear" in the sub

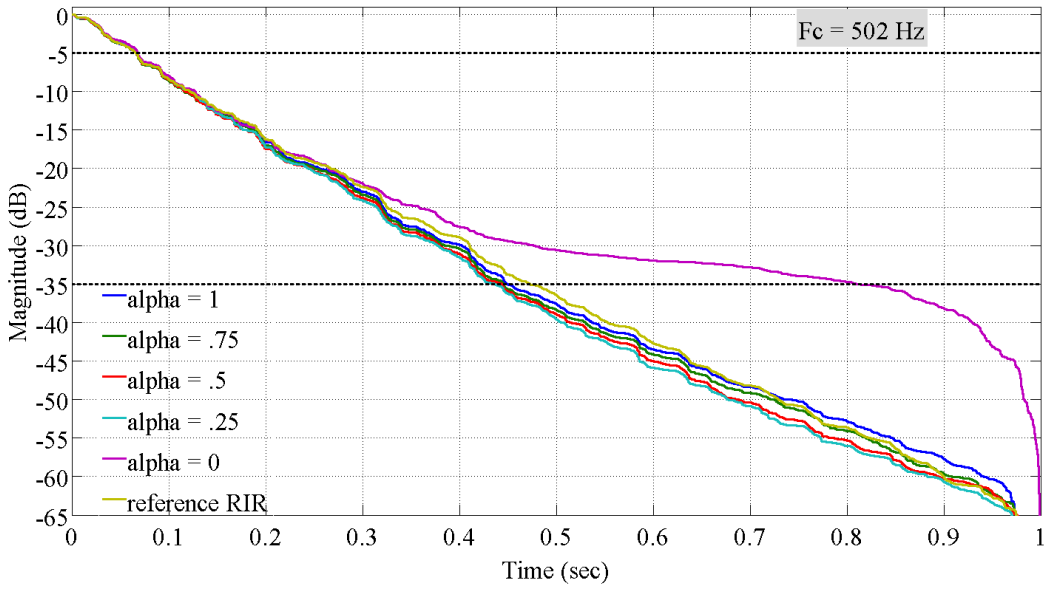
bands that do not contain those frequencies. Consequently, the upper integration limit of the EDC can be extended (up to the level of the background noise).

The sub band centred at 31.5 Hz, for which an upper integration limit of 1.4 seconds has been used, shows that the achieved dynamic range was higher when the presto-chirps stimulus (mainly for α=1) was used, compared with the use of the reference probe (10 seconds long linear chirp, 10 Hz - 5 kHz). The dynamic range was also higher (mainly for α=1) in the 63 Hz sub band. These results justify the fact that the acoustic parameters (Table 8) in the sub bands 31.5 Hz and 63 Hz, are outside the DLs.

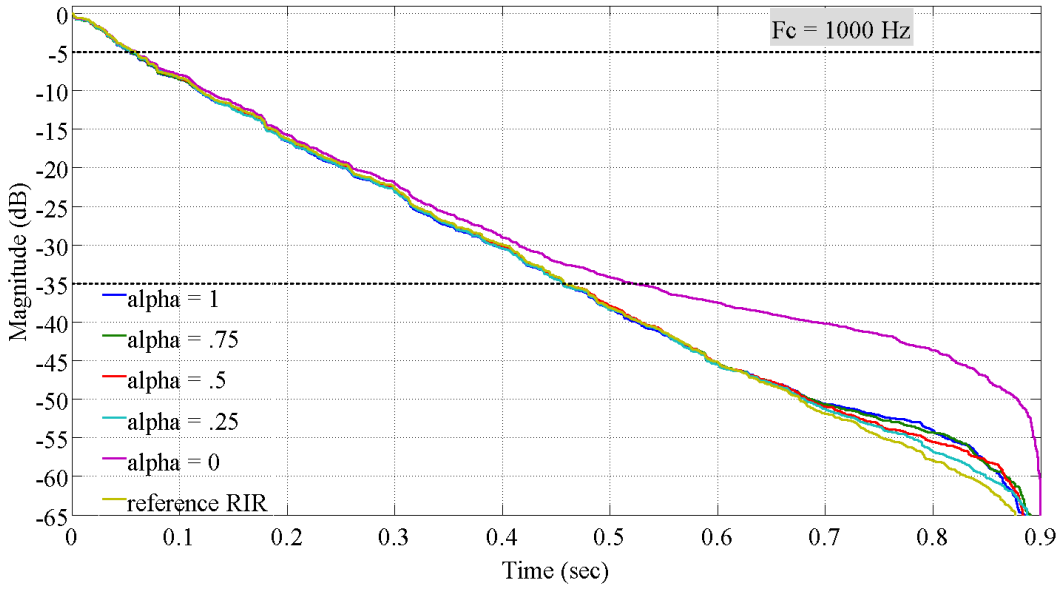



63 Hz
125 Hz
250 Hz
Fc = 64 Hz
Fc = 126 Hz
Fc = 252 Hz
Magnitude (dB)
Time (sec)
alpha = 1
alpha = .75
alpha = .5
alpha = .25
alpha = 0
reference RIR

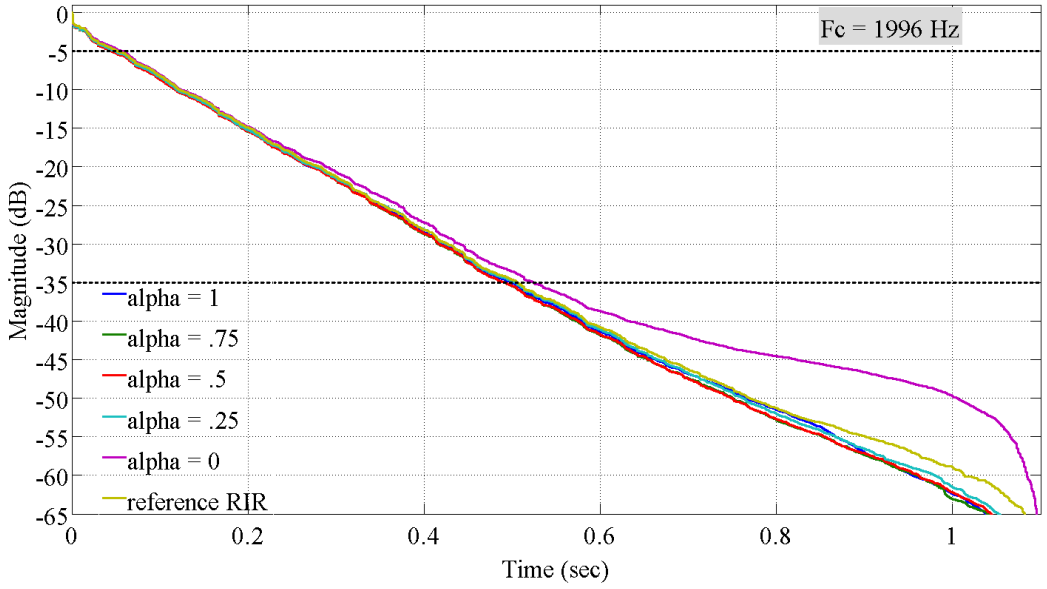
500 Hz



1 kHz



2 kHz



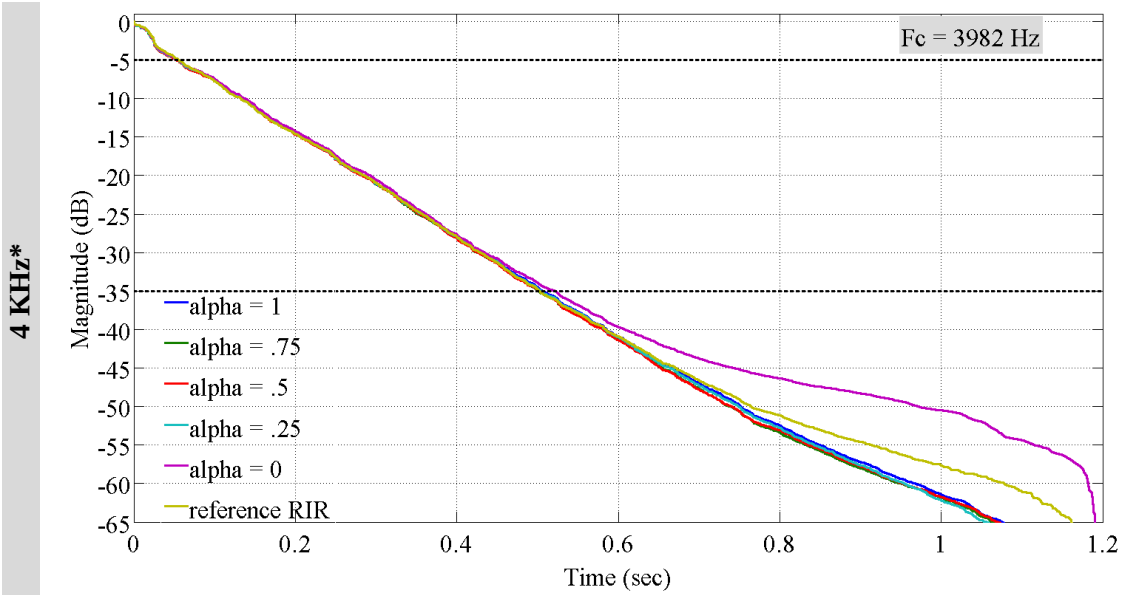


Figure 57 – EDCs of the measured RIRs (as function of α) plus the reference RIR. Each figure accounts for a different sub band as labelled in the left. The dashed lines indicate the decay range generally used to calculate the acoustic parameter RT30. The label 'Fc' represents the central frequency of the octave band filter used in each sub band.

Table 9 reports the SNRs of the measured and reference RIRs, as given by the plug-in "aurora". The SNRs in the sub bands 31.5 Hz and 63 Hz, for the Hanning windowed case (α=1), are about 10 dB higher than the SNR of the reference measure. Such higher SNRs obtained by the presto-chirps stimuli in those sub bands is the result of the averaging effect discussed in Section 4.2.1.5. In fact, as has also been highlighted in Figure 36, the SNR decreases in the lower sub bands as α decreases, which denotes an increased synthesis noise.

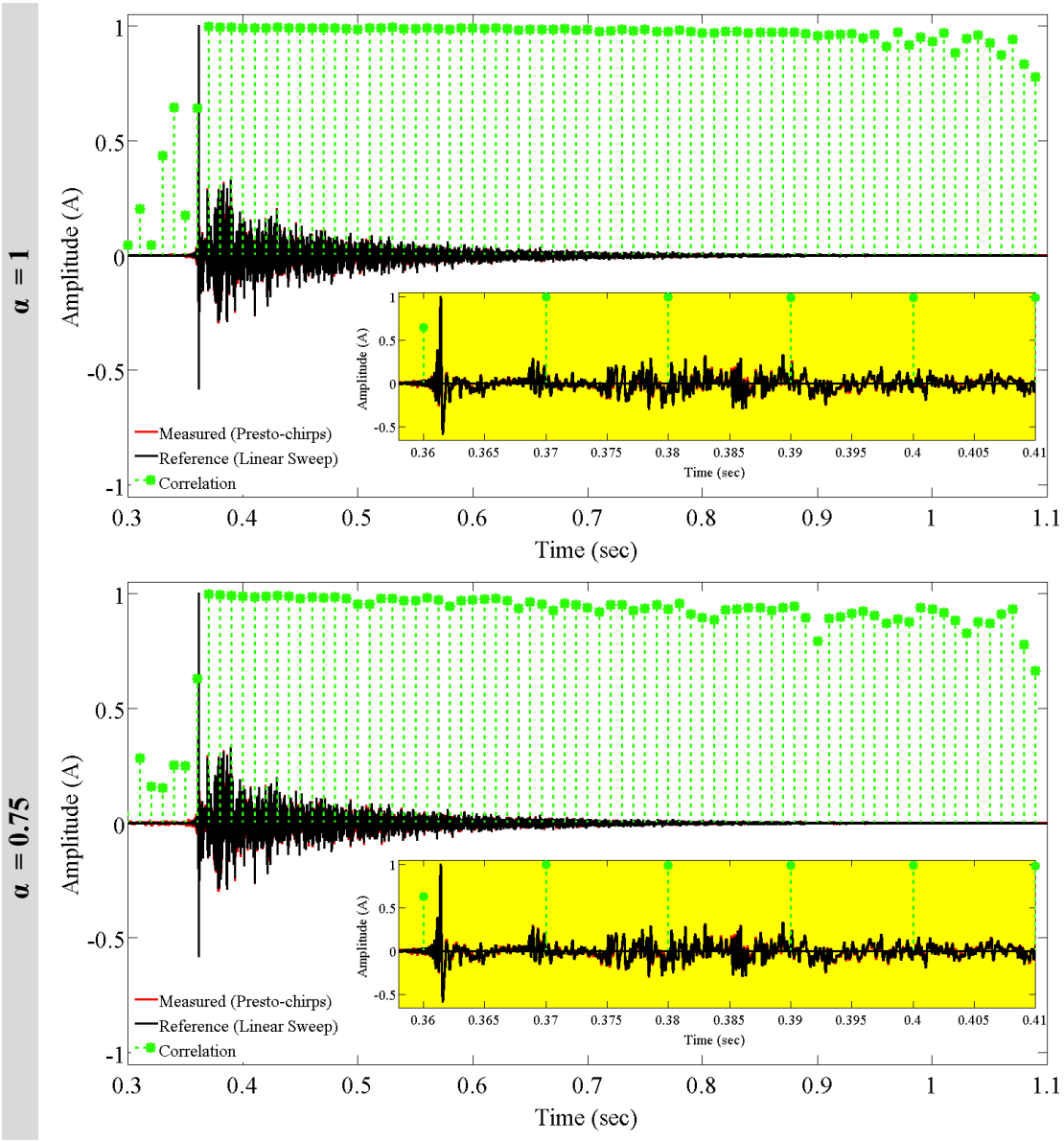
SNR [dB]

Bands [Hz]	31.5	63	125	250	500	1k	2k	4k*
LIN chirp	16.5	22.4	30.1	32.8	37.3	39.8	35.7	35.8
α = 1	26.8	32.8	31.9	33.2	38.4	40.1	36.3	35.9
α = 0.75	18.3	27.5	32.3	34.1	39.2	39.9	36.5	36.3
α = 0.5	14.2	25.9	31.4	33.4	40.0	39.4	36.6	36.4
α = 0.25	13.4	24.7	30.5	33.1	39.8	39.9	35.8	35.9
α = 0	14.4	15.0	24.7	28.2	26.6	32.7	33.3	34.8

Table 9 – SNR given by the software Audacity. The first row, in blue, refers to the reference RIR measured using a 10 seconds long chirp (27 – 4.7 kHz).

4.3.2.2 Time Domain Analysis

The RIR calculated for α=1 is strongly correlated with the reference RIR for most of its length. Correlation slightly decreases at the tail of the RIR, due to a higher background noise. For α=0, the tail of the RIR is poorly correlated, which outlines the increased presence of the synthesis noise. This result has been theoretically demonstrated in the simulated scenario (Section 4.2.2). In the yellow box of the bottom panel, α=0, it is shown magnified the main peak of the RIR.



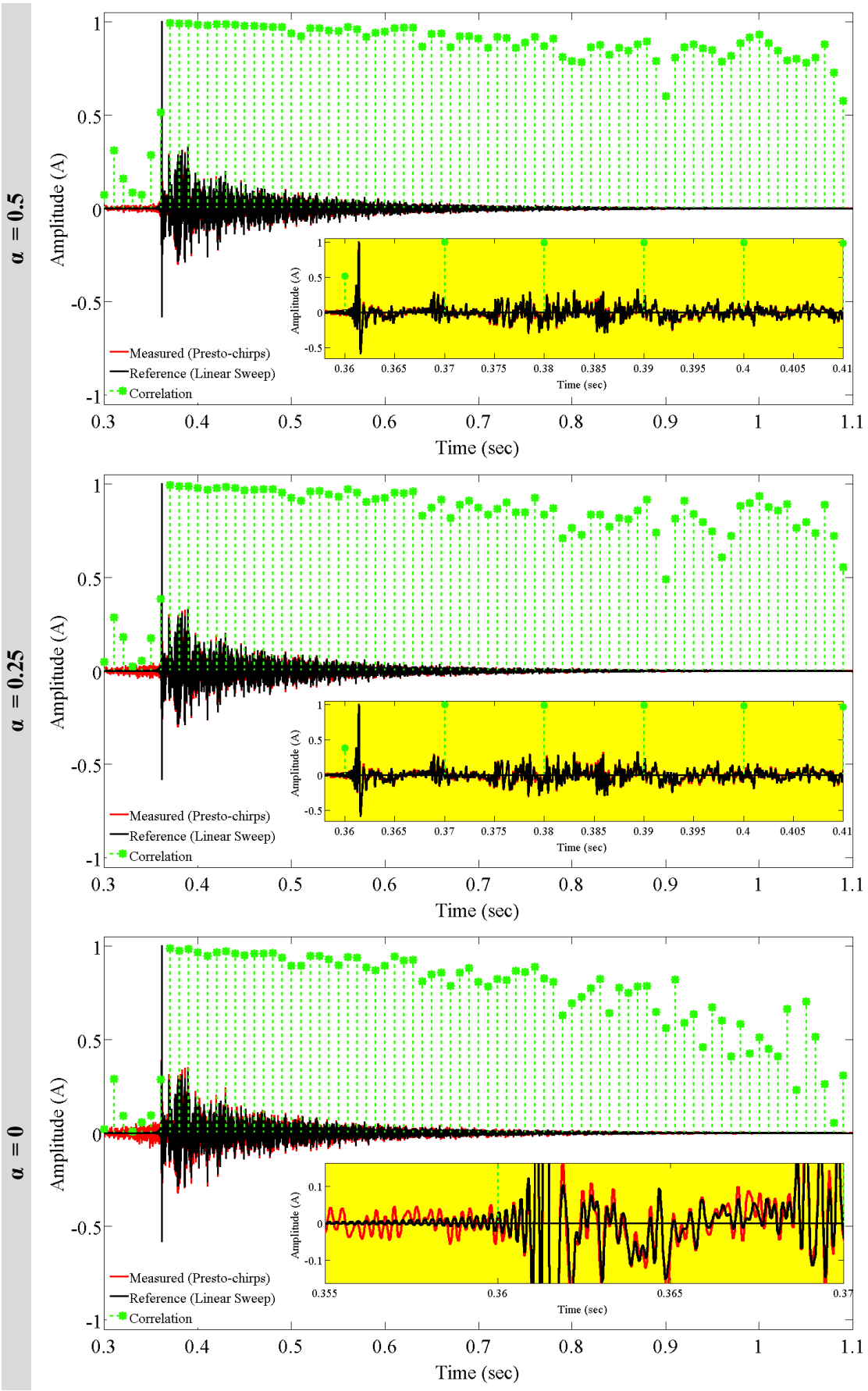

α = 0.5
α = 0.25
α = 0
Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

Figure 58 - Comparisons between the reference RIR (in black) and the measured RIRs (in red). The green marks point out the degree of correlation between 10 ms long blocks of the two RIRs. In the yellow boxes the first 40 ms of the RIRs are shown.

Figure 59 shows the correlations of the whole measured RIRs with the reference one versus the different Tukey windows used. Correlation is higher for α=1 compared with the other windows, and then progressively decreases.

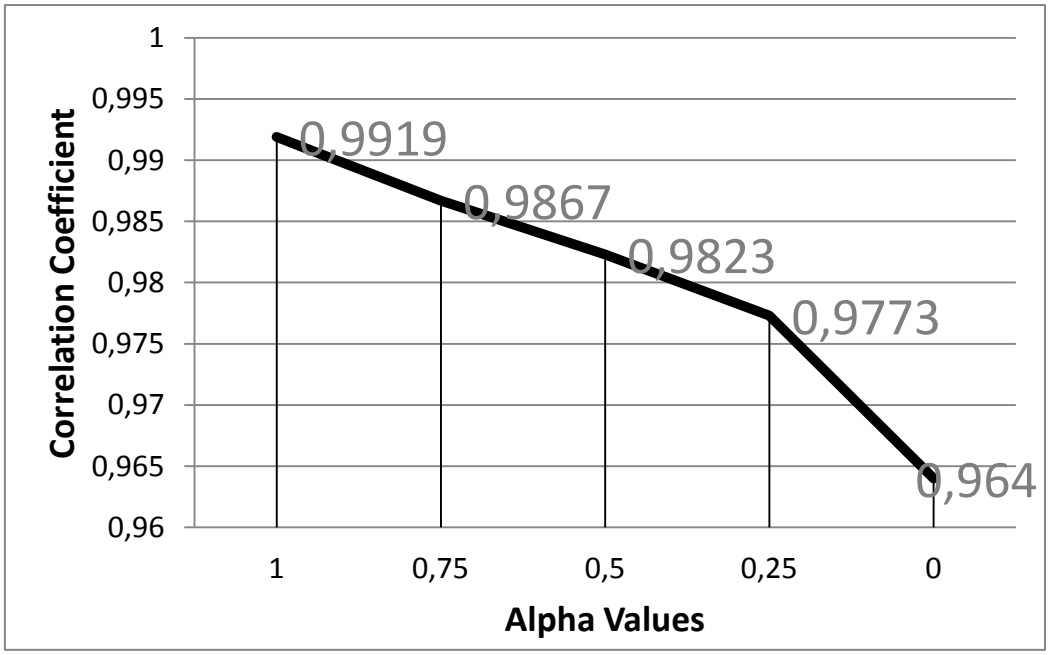
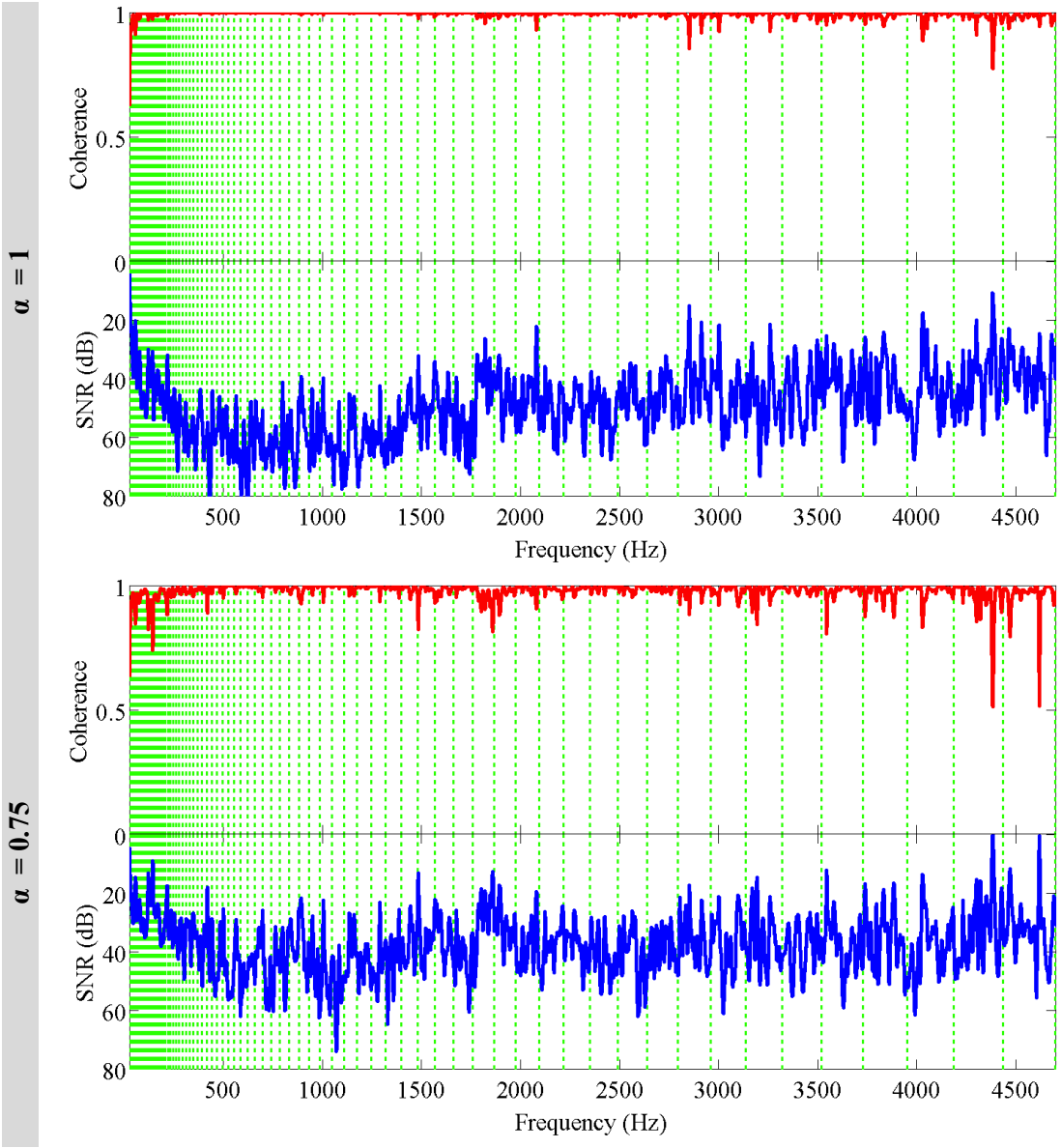


Figure 59 – Degree of correlation between the reference RIR and the RIRs measured with the different α-windowed stimuli.

4.3.2.3 Frequency Domain Analysis

Figure 60 shows the coherence and SNR functions, and points out that the closest match between the reference and the measured FRFs is obtained when Hanning windowed presto-chirps (α=1) are used. The SNR is higher for α=1, especially between 500 Hz and 1 kHz. For all the other tested configurations, the coherence

lowers as α decreases, mainly at low frequencies as a result of unmitigated frequency ripples.



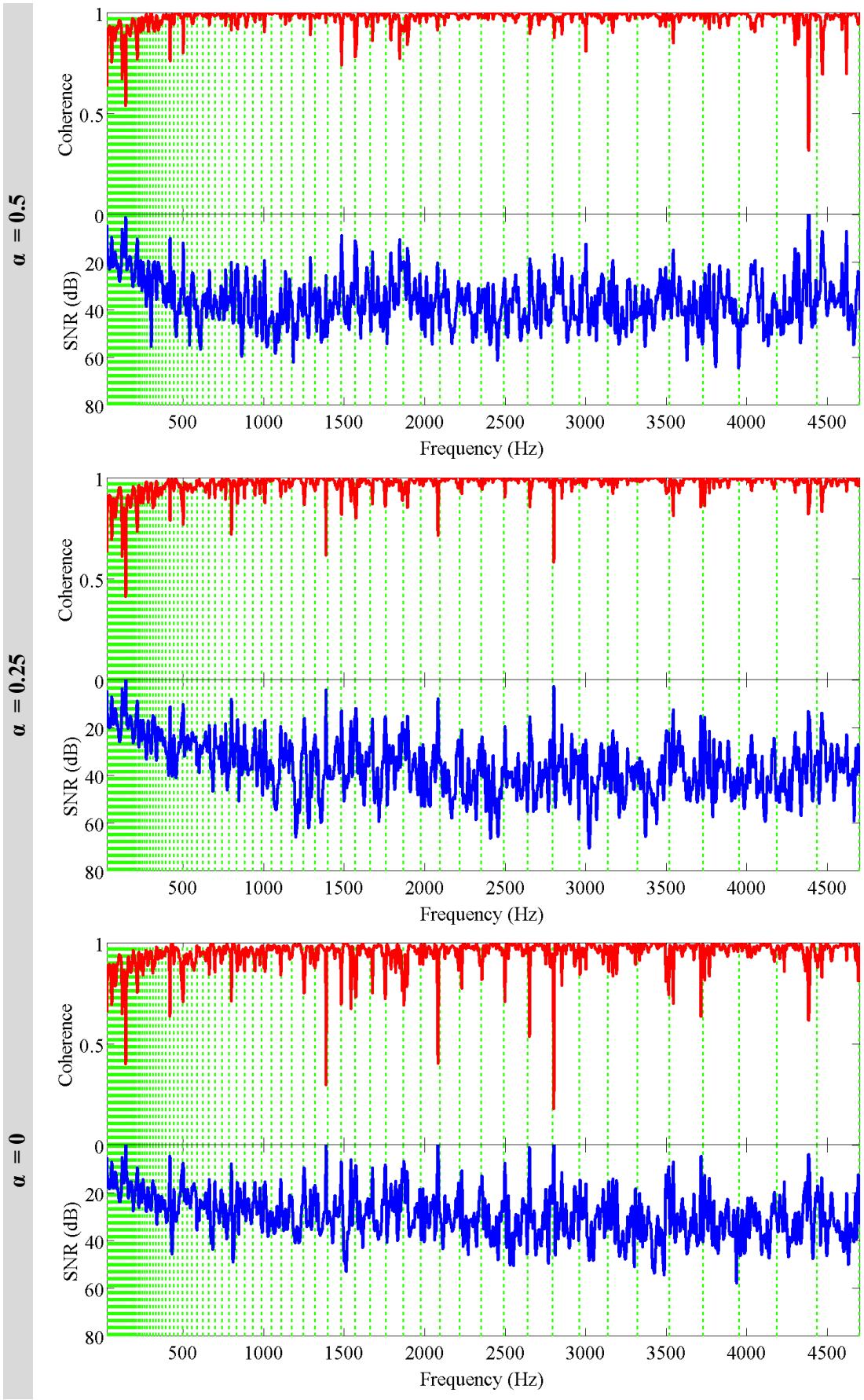


Figure 60 – In red is shown the coherence function, and in blue the SNR, between the reference FRF and the FRF calculated using different α related stimuli, as labelled.

4.3.2.4 Discussions of the Case Study

For the presented validation, it has resulted that using Hanning-windowed presto-chirps results in the most accurate measurement of the RIR compared with the other cases. Such accuracy has been confirmed by the analysis of the acoustic parameters, and the time and the frequency responses. However, the low SNR of the reference RIR, at the sub bands 31.5 Hz and 63 Hz, hindered an accurate evaluation of the acoustical parameters in those bands. Nonetheless, Hanning-windowed presto-chirps provided a greater SNR at the low frequency bands, compared with the reference probe signal (a 10-second long linear chirp), as pointed out in Table 9. This result should be due to the averaging effect discussed in 4.2.1.6, and it reinforces the idea that presto-chirps could be valuable in dealing with high noise sub bands. However, a more detailed evaluation of the accuracy of the method at the very low end of the spectrum should be addressed in future works.

In contrast, all the other tested window configurations have resulted in less accuracy than the $\alpha=1$ case. This is chiefly due to the bias introduced in the measured RIR by the stimuli, especially at the low frequency bands. It can then be concluded that for this tested scenario, only Hanning-windowed presto-chirps could achieve a reliable measurement of the RIR.

Appendix C reports two others in situ measurements. Results for those cases are similar to the ones presented here.

4.4 Discussions on the Presto-chirps Method

Impulse response measurements of several cases studio, using presto-chirps stimuli, have been presented. Firstly, two (software) simulated scenarios were analysed with the aim of highlighting, on one hand, the measurements accuracy obtainable for each tested Tukey windows applied to the presto-chirps. On the other hand, to define how the duration of the presto-chirps influences their spectral content, and consequently their synthesised frequency response.

In addition, the presto-chirps method has been used to measure the RIR of three real venues, using a stimulus obtained from an actual song. The measured RIRs have been compared with a reference RIR measured using a 10 seconds long linear chirp. From these tested cases, it has resulted that using Hanning-windowed presto-chirps leads to the best results, in terms of accuracy / less introduced bias, both in time and in frequency domain, compared with the other tested window configurations. For such a case, the measured acoustic parameters were within difference limens with the reference values. However, this could not have been completely proven in the octave bands centred at 31.5 Hz and 63 Hz, due to the poor SNR ratio of the reference RIR, also because of the inadequate measurements equipment used. Mainly, the loudspeaker did not have sufficient power at low frequencies.

The obtained results confirm that the bias introduced in the RIRs measured using Hanning-windowed presto-chirps stimuli should be unperceivable. For the other tested configurations, the measured acoustic parameters were greater than the difference limen, mainly at low frequency bands. This consequently suggests that

stimuli composed of presto-chirps windowed with Tukey windows different from the Hanning cannot perform accurate acoustic measurements.

For the used measurement level of 70 dB (L_{Aeq}), for both the presto-chirps stimuli and the 10 seconds long linear chirp used to calculate the reference RIRs, it has resulted that the presto-chirps stimuli can generally achieve a higher SNR. The SNR was about 10 dB higher at the lower octave bands (31.5 Hz and 63 Hz) for the Hanning-windowed stimulus compared with the reference RIR (nonetheless, the same measurements equipment was used). The reason for such higher SNR has been attributed to the "averaging" process discussed in Section 4.2.1.5. This, although it should be further assessed, is a positive feature of the presto-chirps method, which partially confirms that the presto-chirps method might effectively be used to tackle the problem of the noisy sub bands, facilitating occupied measurements. Moreover, the highest SNR achievable could be exploited to reduce the level of the measurement probe, which might reduce its obtrusiveness.

The attempt to reduce the annoyance to audiences during acoustics testing, has motivated the use of a music-based stimulus, which has been derived from an actual song composed using all the 88 semitones of a grand piano. The original score, from the Finnish composer Tero-Pekka Henell, has been slightly modified (besides adding the 89^{th} notes) to allow the measurements of up to one second long RIRs. It has been discussed that such limit is inherent to the appearance of residuals in the measured RIR. However, since residuals have limited frequency content, they only affect specific frequencies "zone" (as shown in Appendix A), for a specific length of the

analysis window. This results that when measuring acoustical parameters, a longer analyses window could be use in those sub bands not affected by residuals.

Psychoacoustic assessments of the stimulus, to verify the reduced obtrusiveness compared with broadband chirps, have not been carried out. This is a current limitation of the presented method, which should be addressed in further research. Besides the musicality aspect, presto-chirps present an advantage over broadband probes related to the possibility of arbitrarily defining the duration and amplitude of each presto-chirp of a stimulus. This flexible solution can be exploited to create stimuli that could deal with critical situations with more ease.

5 The Masked-MLS Method

To address the limitations of the presto-chirps method to resolve acoustic measurements in the high end of the audio spectrum, a second measurement method has been devised to measure the audio spectrum from 4.7 kHz to 16 kHz. It has been named the masked-MLS method, because it utilises a (music) masker to mask an MLS probe, which is used for the measurements.

The idea of utilising an MLS to perform occupied measurement is not new. In 1984 Schroeder carried out impulse response measurements of classrooms while he was teaching, using a low level MLS. This was deemed possible thanks to the inherent MLS noise immunity, which allows it to be used with a very low, or even negative, SNR (the MLS is described in Section 2.4.4). However, the use of a low-level MLS might require an over prolonged averaging process to increase the SNR to levels that would permit accurate measurements of RIRs. Unfortunately, thermal fluctuation produce by heating, ventilation, and air conditioning systems, are likely to increase the time variability of venues. For this reasons, prolonged averaging in time variant environments poses difficulties in the achievement of reliable measurements, as the noise rejection property of MLS contrasts with its susceptibility to time variance (Vorlander & Kob, 1997; Svensson & Nielsen, 1999; Satoh et al., 2002). This results in the impossibility of indefinitely increasing the SNR through prolonged synchronous averaging (Serafini et al., 2010), which is a serious limitation for occupied measurements performed using low level MLS.

The masked-MLS method looks at addressing those issues by presenting a measurements framework that can reduce the duration of measurements, and at the same time reducing nuisances to audience. Therefore, the presented method first aims at avoiding time variance drawbacks, while achieving a SNR in compliance with the standard requirements. Secondly, to present the measurement probe in a masked fashion, to allow measurements to be taken in the background, while the audience is present in the venue.

The masked-MLS method tackles both problems by using a low-pass filtered music (fc = 4.7 kHz) to mask the MLS probe used for the measurements. This solution achieves a twofold effect. On the one hand, the use of masker music provides the audience with an entertaining distraction. On the other hand, the forward masking induced by the music is exploited to mask a medium level MLS, which consequently limits its obtrusiveness. Low pass filtering the music is needed to ensure that the MLS probe will only deal with the background noise, and so enabling shorter measurements. To exploit properly the masking from music, the MLS is frequency shaped according to a masking threshold derived from the music, using a psychoacoustic model.

5.1 Theoretical Background

This paragraph reviews some relevant psychoacoustics facts that form the theoretical background of the masked-MLS method.

5.1.1 Threshold of Hearing and Equal-loudness-level Contours

The first topic to be presented when talking about psychoacoustics is the absolute threshold of hearing (ATH), which is also called the threshold in quiet (THQ). The ATH is the minimum sound pressure level of a tone that the human auditory system can perceive. Terhardt (1979), proposed a mathematical expression to represent the ATH, which is

$$T_q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\left(\frac{f}{1000}\right)-3.3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 \qquad (60)$$

where $T_q(f)$ is the threshold of hearing in dB and f is the frequency in Hertz. The second topic is related to the equal-loudness-level contours (ELLCs), first investigated by Fletcher & Munson (1933), which represent the loudness perception versus the frequency of a pure tone under free-field listening conditions. Figure 61 shows the ELLCs, redefined since the first version proposed by Fletcher & Munson (1933), extracted from the standard BS EN ISO 226:2003. The solid lines represent the perceived loudness (in phon) of normal hearing people when listening to a tone stimulus with a certain level and frequency. The threshold of hearing is displayed with a dashed line, whereas the dotted lines at 10 and 100 phon indicate an unconfirmed trend due to a lack of experimental data. The phon scale is a subjective scale of

loudness, calculated comparing the loudness sensation of a reference 1 kHz tone versus the same loudness perceived at other frequencies. The phon scale at 1 kHz intersects, by definition, the corresponding dB (SPL) value. The hearing sensitivity peaks in the frequency range from 2 to 5 kHz (due to the resonance of the ear canal), and it rapidly drops for frequencies above 5 kHz.

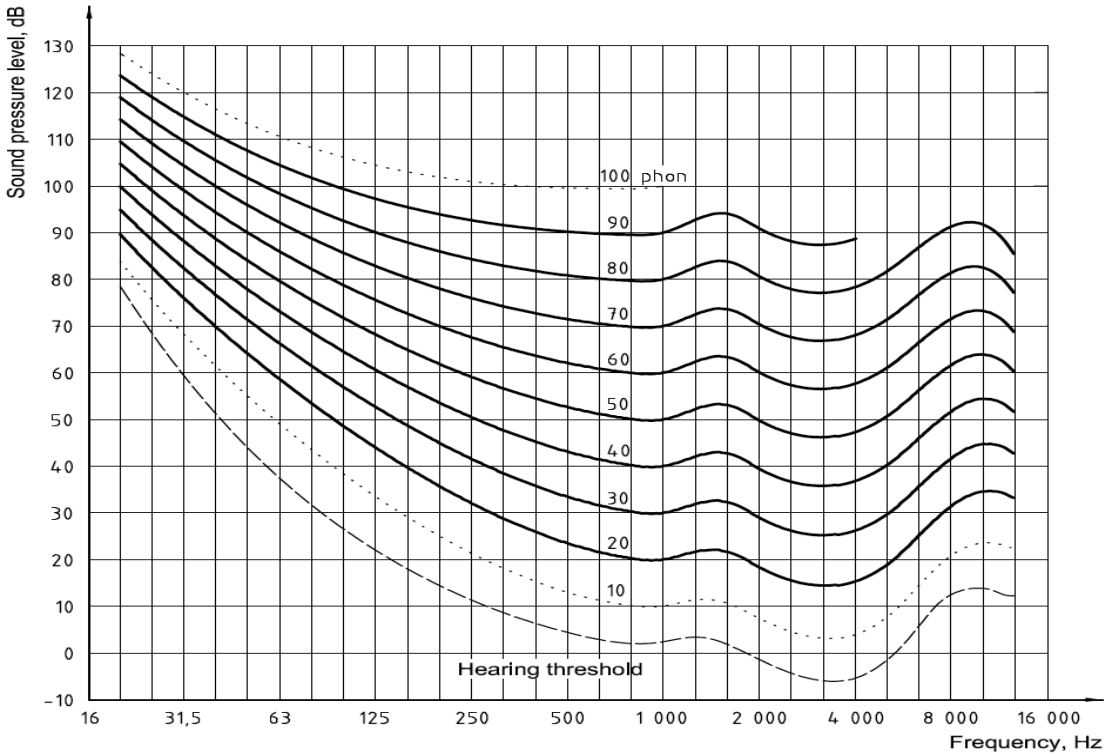


Figure 61 – Equal-loudness-level contours as published in the BS EN ISO 226:2003.

5.1.2 Auditory Filter Bank and the Bark Scale

Masking is a psycho-physiological effect due to the spread and overlapping of sound waves into the cochlea. Fletcher (1940) highlighted that a frequency-to-place transformation happens in "characteristic zones" of the basilar membrane, each having a specific characteristic frequency (CF), and "critical" bandwidth (CB). He suggested that the auditory system could be modelled through a set of band pass

filters, which all together constitute the auditory filter bank. Since him, several authors have developed and used methods to measure more finely the shapes and the widths of the auditory filter bank. In fact, during the years the shape of the auditory filters has been subjected to continued revisions, as other measurements techniques were available (e.g., direct measurements of masking effects, and loudness comparisons). The definition of the shapes of the auditory filters and the developing of models that best fit experimental data is an active research field in acoustics. A detailed description of the recent developments of auditory filter models and their applications can be found in Lyon (2010).

These efforts led to the definition of an auditory filter bank subdivided into the 24 critical bands (or channels). Table 10 reports the CFs and CBs of each channel (from Fastl & Zwicker, 2007). The cut off frequency of the 24th channel has been extracted from Smith III & Abel, (1999).

Channel	CF (Hz)	Range (Hz-Hz)	CB (Hz)
0	50	0-100	100
1	150	100-200	100
2	250	200-300	100
3	350	300-400	100
4	450	400-510	110
5	570	510-630	120
6	700	630-770	140
7	840	770-920	150
8	1000	920-1080	160
9	1170	1080-1270	190
10	1370	1270-1480	210
11	1600	1480-1720	240
12	1850	1720-2000	280
13	2150	2000-2320	320
14	2500	2320-2700	380
15	2900	2700-3150	450
16	3400	3150-3700	550

17	4000	3700-4400	700
18	4800	4400-5300	900
19	5800	5300-6400	1100
20	7000	6400-7700	1300
21	8500	7700-9500	1800
22	10500	9500-12000	2500
23	13500	12000-15500	3500
24	18000	15500-20500	5000

Table 10 – Auditory filter bank: channels centre frequency and bandwidth.

In the 1961, Zwicker proposed the use of a psychoacoustic scale, which is, as he defined it "…a frequency scale on which equal distances correspond with perceptually equal distances". The bark scale is related to the critical bands, and has been named after H. Barkhausen who first carried out measurements of loudness. In practice, in the Bark scale, the distance of one bark (z) represents the distance of one critical band. Zwicker & Terhardt (1980) proposed a function, graphically shown in Figure 62, which can be used to convert frequencies (in Hertz) into a Bark quantity.

$$Z(f) = 13 \cdot \tan^{-1}(0.76 \cdot 10^{-3} \cdot f) + 3.5 \cdot \tan^{-1}[(f/7500)^2] \qquad (61)$$

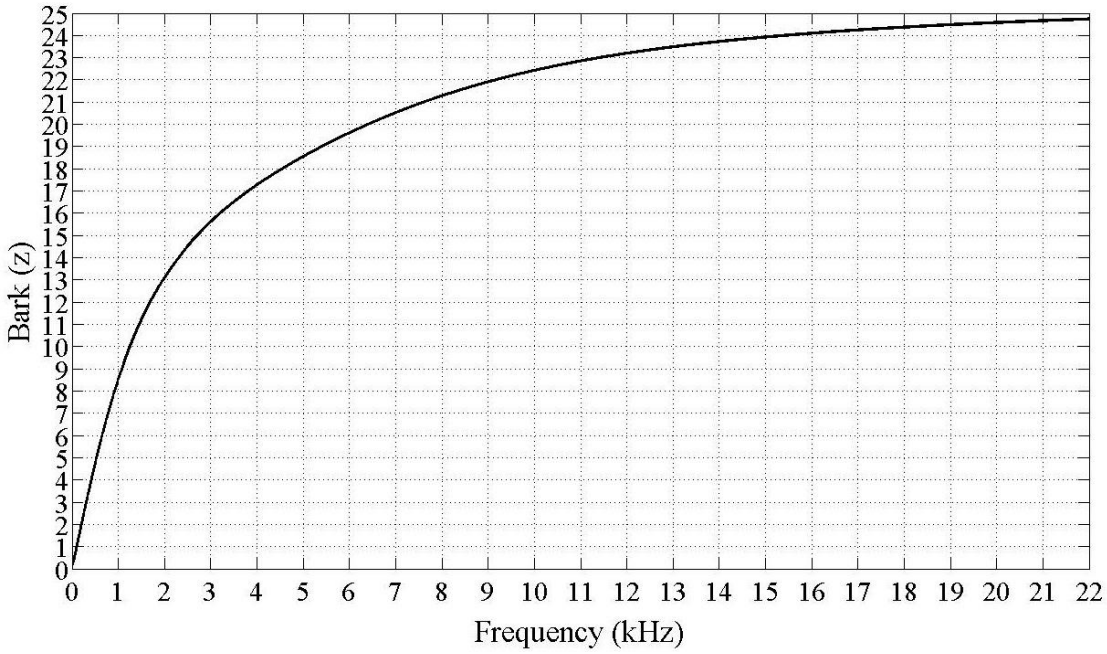


Figure 62 – Relation between frequency and bark.

5.1.3 Masking Effects

A masker sound creates a frequency-localised increase of the threshold of hearing, which is generally referred to as the masking threshold. In other words, masking is related to the increase of the threshold of hearing of one sound (the maskee) caused by the presence of another sound (the masker). Any sound, be it formed by a single tone or several tones (complex sound), induce masking. However, the masking threshold that such sound generates varies based on its spectral characteristics. Tones, and complex sounds produce different masking thresholds, as discussed in Section 5.1.5. The ability of a sound to mask another sound, when both are simultaneously analysed by the auditory filters, depends on the frequencies of the masker and of the maskee. Masking can also be produced when the masker and the maskee are not simultaneous events, but there is some time gap between the two signals. In this case,

the phenomenon is called temporal masking, and its amount depends both on the levels, and on the time gap between the two sounds.

To recap, masking is composed of two distinct phenomena, namely simultaneous and temporal masking:

- Simultaneous masking is a frequency domain phenomenon (and hence it is called spectral masking), which refers to a situation when both the masker and the masked sounds are simultaneously analysed by the auditory filters. The spectral distribution of the masker determines which frequencies will be masked and to what extent (level).

- Temporal masking is a time domain phenomenon associated with a masker that masks a following (post-masking) or a preceding (pre-masking) signal. There is a marked difference between the two effects with pre-masking lasting for about 10 ms, and post-masking lasting for about 200 ms.

Figure 63 reports a schematic representation of both the temporal masking effects (pre and post), and of the simultaneous masking effect. These for a 200 ms long white noise masker - adapted from Painter & Spanias (2000).

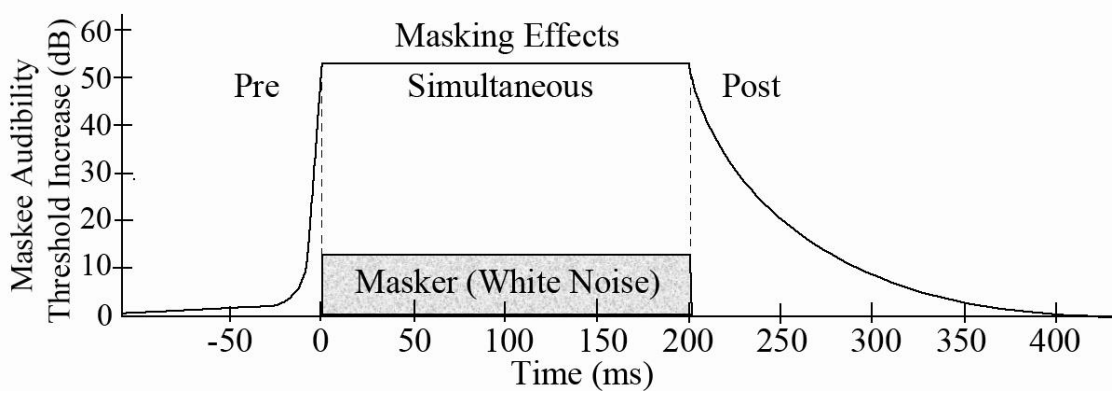


Figure 63 – Temporal masking for a 200 ms long white noise masker.

5.1.4 The Spread of Masking

The masked-MLS is based upon an effect called the forward spread of masking. To better visualise this phenomenon, Figure 64 reports the masking thresholds produced by a critical-band width noise, for the labelled central frequencies, related to the detection of a tone signal. Masking thresholds are high when the frequencies of the masker and the maskee are similar, but it also extends towards neighbouring critical bands. Such effect is what is called the "upward" spread of masking (or sometimes inter-band masking). It is opposed to the "backward" spread of masking, which involves the masking effect towards lower frequencies. Differences between the two effects are linked to the physiology of the cochlea. In fact, as first demonstrated by Bekesy, sound waves with higher frequency travels longer towards the apex of the basilar membrane compared with low frequency sounds. The consequence of this effect is that, generally, high frequencies sounds can be masked more easily than low frequencies sounds.

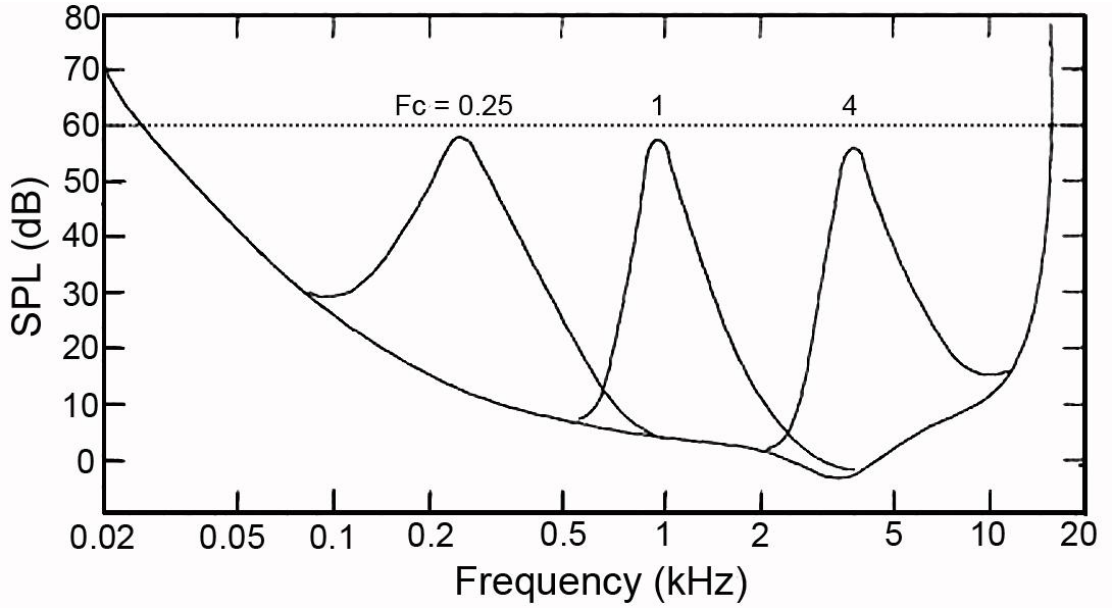


Figure 64 – Level of a test tone just masked by a critical-band width noise centred at 0.25, 1 and 4 kHz (adapted from E. Zwicker and H. Fastl, Psychoacoustics—Facts and Models (Springer-Verlag, Berlin, Germany, 1990).

Figure 65 shows that by using the Bark scale in place of linear frequency, masking thresholds exhibit symmetrical shapes, regardless of their central frequency. Masking thresholds are shown as a function of the frequency, and of the level of a just perceivable test tone masked by a one bark-width noise of 60 dB SPL. The noise was centred at the frequencies labelled in the figure. The dashed line represents the ATH.

Human auditory system is non-linear, in which the perception strongly depends on the level. The masking thresholds consequently present a non-linear behaviour. Figure 66, adapted from Bosi (2002), shows the excitation thresholds due to a masker noise (one critical band width centred at 1 kHz / 8 bark) for several masker levels. It can be seen that the masking thresholds become shallower towards the high critical bands as the level of the stimulus increases.

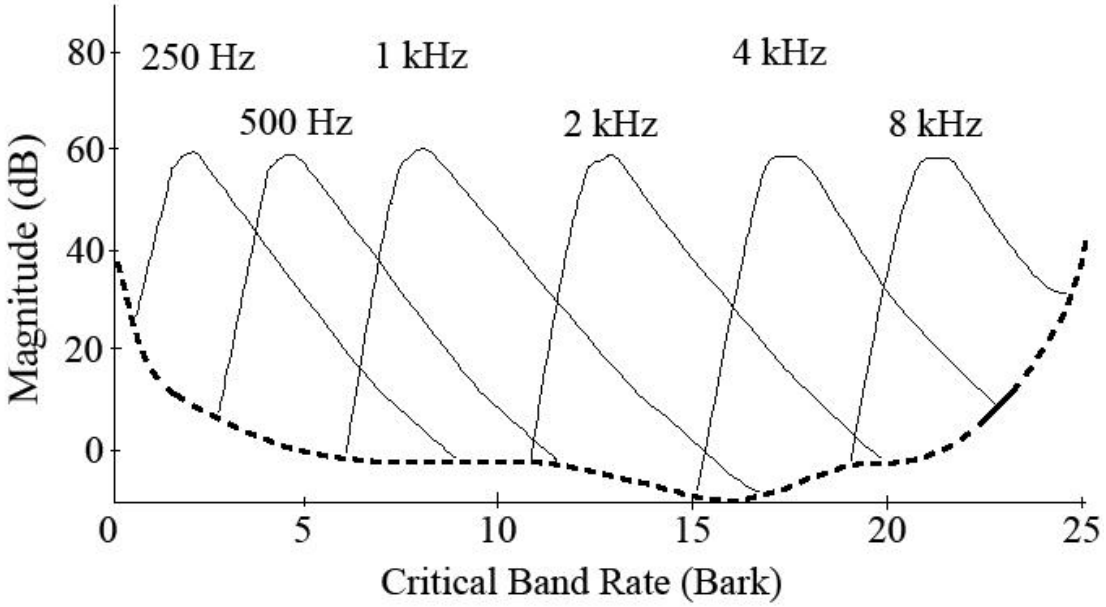


Figure 65 – Masking thresholds for different central frequencies in the Bark scale (adapted from Fastl, 2013).

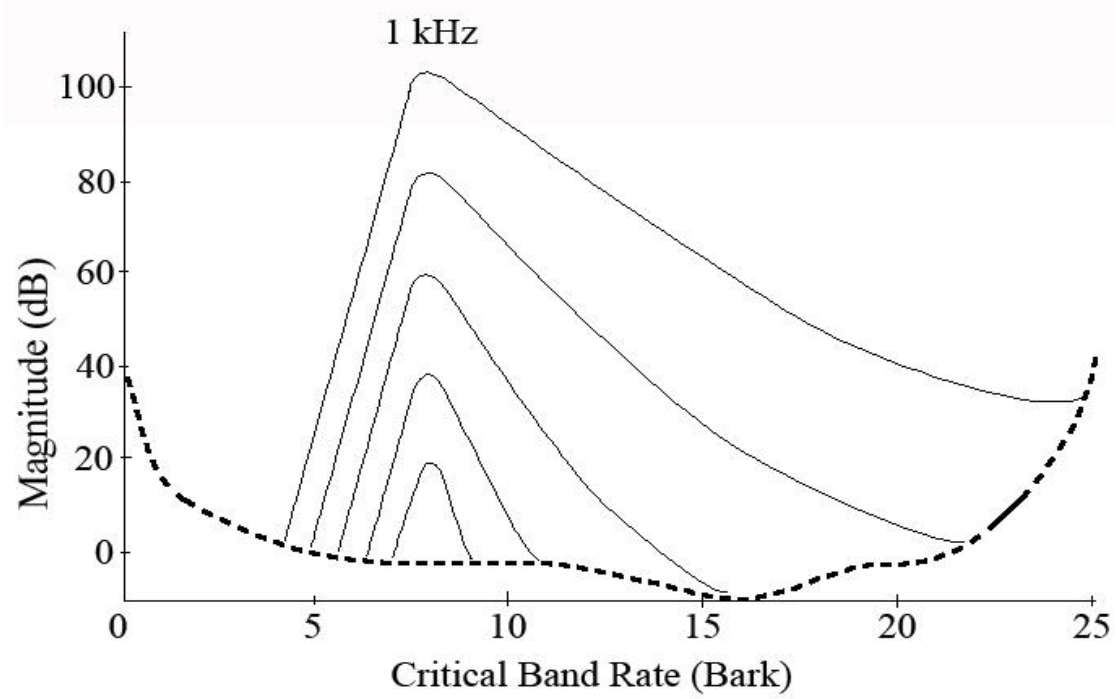


Figure 66 – Masking thresholds for various levels of a critical bandwidth masker noise centred at 1 kHz. The dashed line represents the ATH.

5.1.5 Asymmetry of Masking

Masking thresholds are functions of the level and the frequency of the masker sounds. A masker is generally termed as tone-like or noise-like, with the latter being a more effective masker. This different masking ability between the two types of maskers is referred as "asymmetry of masking". Figure 67 and Figure 68 reports (adapted from Fastl & Zwicker, 2007) the masking thresholds of a tone maskee masked by a critical-band width noise and a tone maskers, respectively. Both maskers are centred at 1 kHz, and thresholds are given for the SPLs labelled in the figures.

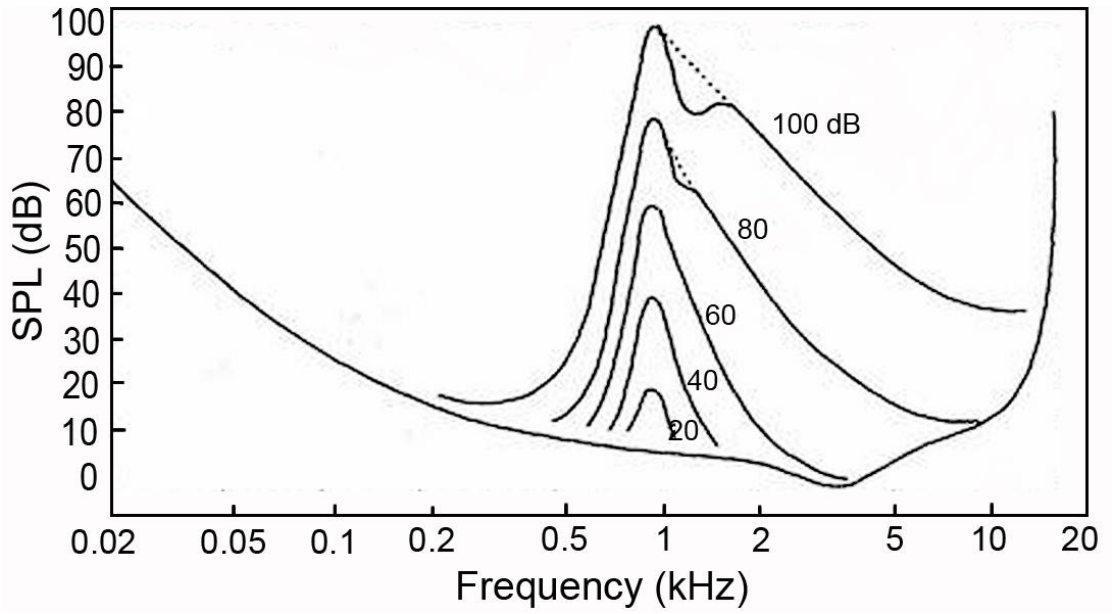


Figure 67 – Level of a test tone just masked by a critical-band width noise for different levels and centred at 1 kHz (adapted from E. Zwicker and H. Fastl, Psychoacoustics—Facts and Models (Springer-Verlag, Berlin, Germany, 1990).

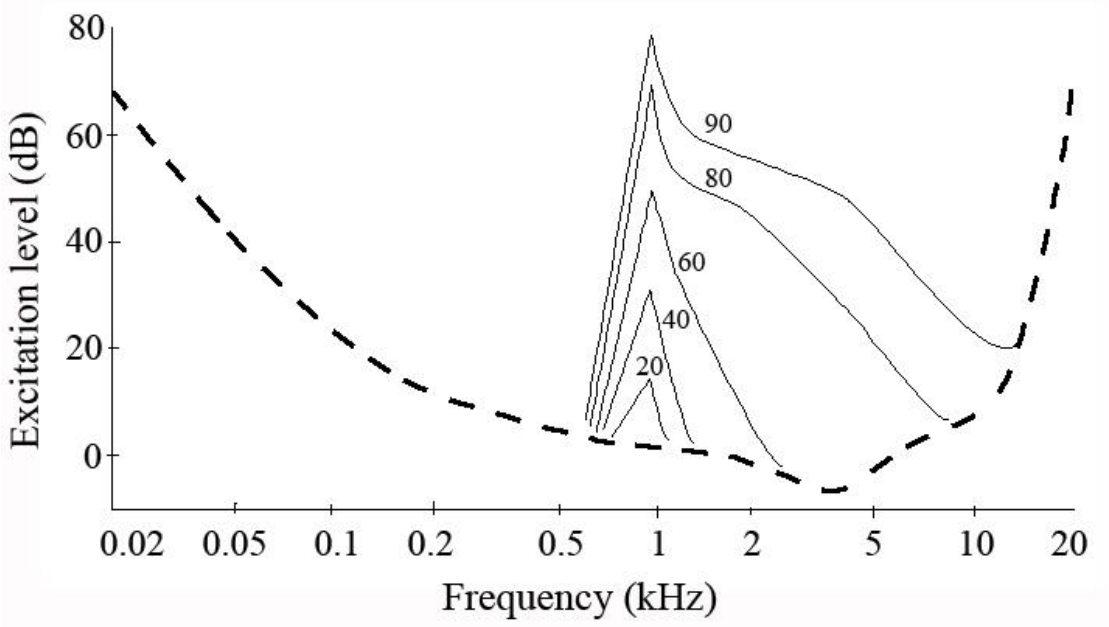


Figure 68 – Masking thresholds for a 1 kHz tone for several SPL levels (as labelled).

It should be noted that there is a marked difference between the two cases. In the first case, a 100 dB SPL noise masker masks a tone (with same central frequency of the noise) with a level few dB lower. In the second case, a 90 dB SPL tone masker masks a tone (with same frequency of the tone) with a level of about 10 dB lower. The ratio between the level of the masker and the level of a just-not-noticeable masked sound is called the signal to mask ratio (SMR). It differs for any combination of masker – maskee. Besides the aforementioned combinations of a noise masking a tone (NMT) and a tone masking a tone (TMT), others are possible. For example, a tone masking a noise (TMN), and a noise masking a noise (NMN). Several researches have been carried out on the various combination of masker – maskee, Egan & Hake, (1950), Hall, (1997), Fastl & Zwicker (2007), and Hellman, (1972), to cite a few. However, the first two cases of the NMT and TMT have been more clearly studied and understood. In contrast, the TMN and the NMN have not been researched as fully, potentially due to the difficulty to discern, in listening tests, the masker from the maskee. Even more complex is the situation when a masker is a complex stimulus

formed by any combination of tone-like maskers. This scenario, which in practice can be related to musical signals, has been investigated by Gockel et al., (2002; 2003). They have shown that the masking produced by complex stimuli also depends on the phases of and the envelope of the tones.

5.2 Implementation of the Method

To allow the functioning of the masked-MLS method a series of steps have to be taken. First, suitable masker music has to be chosen, and a masker threshold calculated from it. Such masking threshold is predicted by means of the psychoacoustic model used in the "MPEG-1 layer 1" codec. Secondly, a linear phase FIR filter is created based on the shape of the masking threshold. An inverse filter is also created based on the former filter. Both filters are used within the measurements chain. The levels of the music and of the MLS are in situ calibrated, to make them correspond, to 75 dB L_{Aeq}, for what concerns the music; to the level predicted by the psychoacoustic model, for what concerns the MLS.

To summarise, the measurement scheme is:

- Finding a suitable masker music
- Low pass filtering the masker music (plus some equalisation and compression)
- Estimation of the "lower" masking threshold from the masker music (by means of the psychoacoustic model used in "MPEG-1 layer 1 codec)
- Definition of an FIR filter and of its inverse, based on the estimated masking threshold (both filters are applied to the MLS during the measurements)
- In situ calibration of the levels of the masker music and the MLS
- Measurements and analysis of the results

5.2.1 Psychoacoustic Analysis of Maskers

The masked-MLS method bases its theoretical ground upon the level-depended spread of masking, since the basic idea behind the method is that an "efficient" masker played at high level should be able to mask a low-level MLS probe. In the last decades, the knowledge of masking effects has been used in conjunction with more refined signal processing techniques to develop "perceptual" audio codec. Such codecs exploit the functioning of human perception to decrease the amount of data needed to represent an audio source, by identifying those frequencies that are inaudible due to masking effects. A well-known and broadly used perceptual codec has been released by the moving picture experts group (MPEG), founded in 1988. The first release was the MPEG-1 layer 1 codec (MP1), in the 1993, followed shortly after by the MPEG-1 layer 2 (MP2), and then, few years later, by the MPEG-1 layer 3 (MP3). This latter determined the popularity of this family of codec. Details of all the three layers can be found in the published standard ISO/IEC 11172-3.

The masked-MLS method has been linked to the psychoacoustic model worn in the MP1 codec. It is used to determine the masking threshold produced by the low pass filtered masker music. This despite it is the most advanced psychoacoustic model available at present days. Nonetheless, it has been preferred over the newest versions because, firstly the codec's algorithms were readily available within the programmers community (due also to the fact that its copyright has expired); secondly, because only the estimation of masking thresholds derived from the masker music are of any interest for the masked-MLS method, and not, for example, the possibility of achieving greater compression rates. For this latter reason, the description of the

architectures of the MP1 codec is not fully covered in this thesis, but is only discussed in relation to those algorithms actually used to calculate the "lower" masking threshold, from the masker music. More specific details of the codec can be found in Pan (1995), Painter & Spanias (2000), and Bosi (2002).

5.2.2 Determination of the "Lowest" Masking Threshold

The MP1 codec uses a 32 (uniformly spaced) bands filter bank to analyse 512 samples long blocks frames of an audio file. For a sampling frequency of 44.1 kHz, each block is 11.6 ms, which gives a frequency resolution of 86.1 Hz/bin. This subdivision is a convenient trade-off for computation complexity and frequency resolution. Moreover, short frames guarantee low latency, which is an important issue in real time applications. To normalise the level of a frame of audio, the psychoacoustic model implemented in the MP1 codec associate a 0 dBFS tone to a 90 dB SPL tone (Painter (2000). This value is related to the dynamic range of a 16-bit format audio wave (which is lesser than the theoretical 96 dB limit due to practical hardware limitations, and of the addition of dither noise). Subsequently, a "tonality" check is carried out for each sub band to find those components of the audio data that have a tone-like or a noise-like behaviour (this is done through an analysis of the magnitude variations ascending-descending of each FFT bin with neighbouring bins). Then, associated to each component, after a process of decimation that reduces their overall number, the following level-dependent spreading function (plus an offset factor discussed later)

$$SF(i,j) = \begin{cases} 17\Delta z - 0.4L_M(j) + 11, & -3 \le \Delta z < -1 \\ (0.4L_M(j) + 6)\Delta z, & -1 \le \Delta z < -0 \\ -17\Delta z, & 0 \le \Delta z < 1 \\ (0.15L_M(j) - 17)\Delta z - 0.15L_M(j), & 1 \le \Delta z < 8 \end{cases} \quad (62)$$

where L_M is the SPL of the masker at bin j and Δz is the difference between the maskee and the masker, thus: $\Delta z = z(i) - z(j)$. The terms z(j) and z(i) represent the critical bands associated with the masker bin j and the maskee bin i, respectively. Figure 69 reports the spreading function $SF(i,j)$ centred at critical band 18 for different levels of masker signals, as function of the bark. The levels are labelled in the figure and the dashed line represents the ATH. Note that, in this model, the spreading functions are independent of the critical bands.

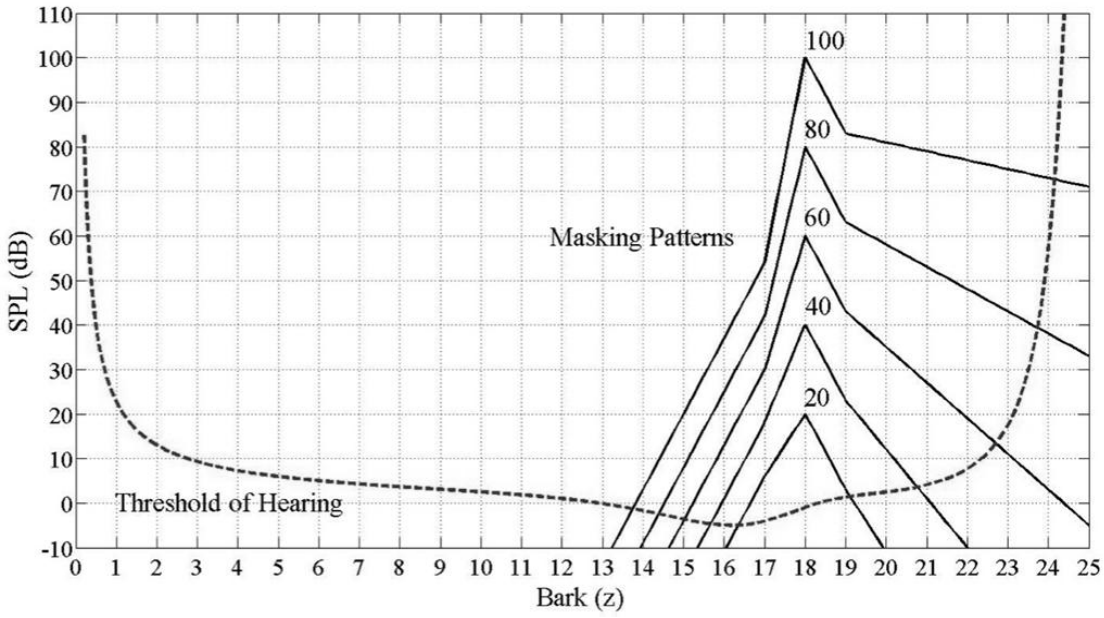


Figure 69 – Spreading functions used in the psychoacoustic model implemented in the MP1 codec.

Although The MP1 uses the same spreading function (Eq. (62)) for both the tone and noise like bins, it offsets the spreading functions, by a specific amount (the SMR)

related to the nature (tone or noise) of the masker. E.g., for a tone–like bin $T_{TM}(i,j)$ its associated spreading function becomes

$$T_{TM}(i,j) = P_{TM}(j) - 0.275 \cdot Z(j) + SF_T(i,j) - 6.025 \quad (63)$$

where $P_{TM}(j)$ is the level in dB of the signal at bin j related to bark $Z(j)$, and $SF(i,j)$ is the spreading function. Conversely, the spreading function associated to a noise-like bin $T_{NM}(i,j)$ is

$$T_{NM}(i,j) = P_{NM}(j) - 0.175 \cdot Z(j) + SF_N(i,j) - 2.025 \quad (64)$$

where $P_{NM}(j)$ is the level in dB of the signal at bin j related to bark $Z(j)$, and $SF(i,j)$ is the spreading function. Figure 70 reports the spreading functions resulting from Eq. (63) and Eq. (64), along with the non-offset spreading function (solid line).

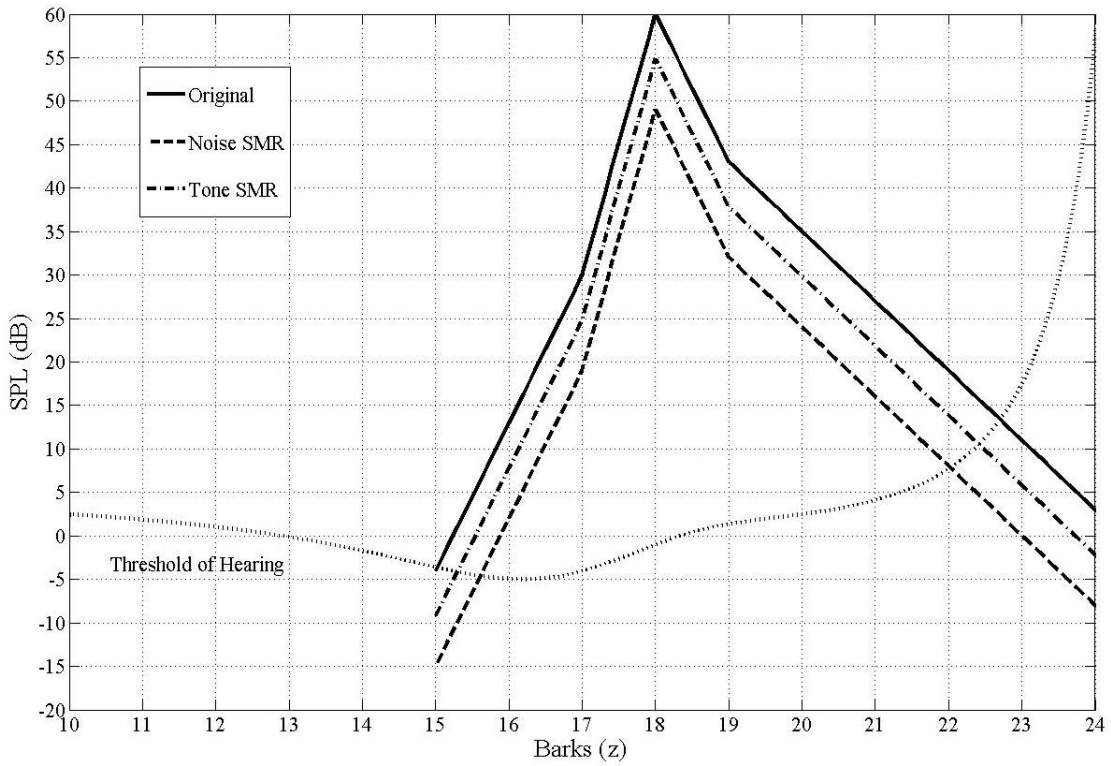


Figure 70 – Spreading functions associated with a tone masker (Tone SMR) and a noise masker (Noise SMR), as used in the psychoacoustic model implemented in the MP1 codec.

Figure 71 shows an example of the analysis of a frame of audio. The solid blue line indicates the frequency response of the analysed audio frame. The red and green circles indicate the tone-like and the noise-like components, respectively. The green and red dashed lines indicate their associated masking thresholds. The green solid line, at the bottom of the figure, represents the threshold of hearing. The dashed black line represents the global masking threshold, which is formed from the combination of all the masking thresholds. Note that the abscissa represents frequency index points, which indicate the FFT bins of the transformed frame of audio.

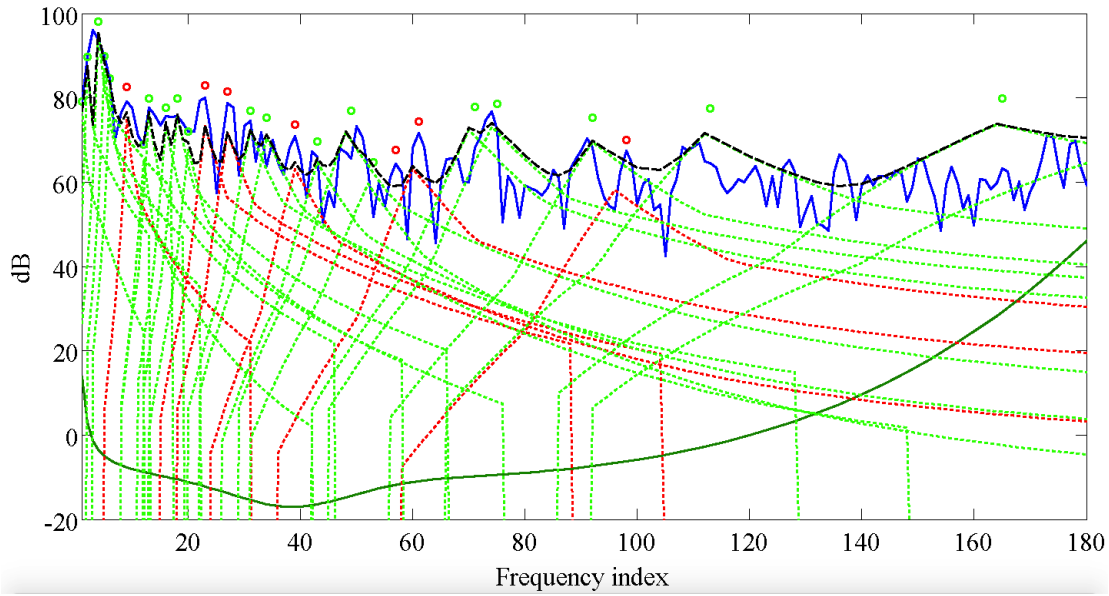


Figure 71 – Making analysis of a frame of audio music. The crosses represent the tonal maskers (TM), and the circles the noise maskers (NM). The ATH is represented with a red dashed line, the average SPL of the frame with a green dotted line, and the global masking threshold by the solid black line.

For each frame of audio, a global masking threshold is calculated and stored. At the end of the analysis process, they are compared, and a "lowest" masking threshold is constructed by taking the lower (local) values among all the thresholds. Figure 72 shows the lowest masking threshold (dubbed "Lowest MP") calculated from the masker music used in the validations of the method. The green line represents the median SPL of the music, and the dashed red line the ATH.

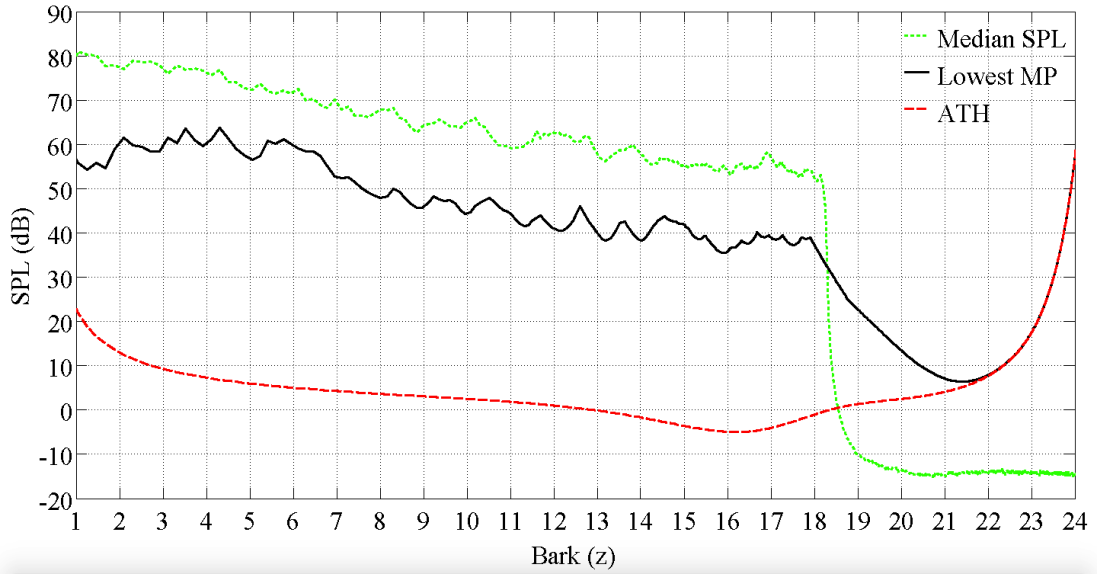


Figure 72 – Masking analysis of the whole masker. ATH = absolute threshold of hearing.

5.2.3 Creation of the Forward and Inverse Filters

A linear phase (FIR) filter is created upon the magnitude response of the lowest masking threshold, from 4.7 kHz upwards (thus from bark 18 to 24). However, since the background noise decreases at higher frequencies (it generally has a pink or brown PSD), for practicality it has been preferred to "flatten" the magnitude response of the filter when the threshold reaches its lower value (i.e. around 8 kHz). Nonetheless, the threshold of hearing increases towards higher frequencies. Figure 73 reports the normalised magnitude response of the FIR filter constructed upon the lowest masking threshold, flattened from 8 kHz to 22 kHz.

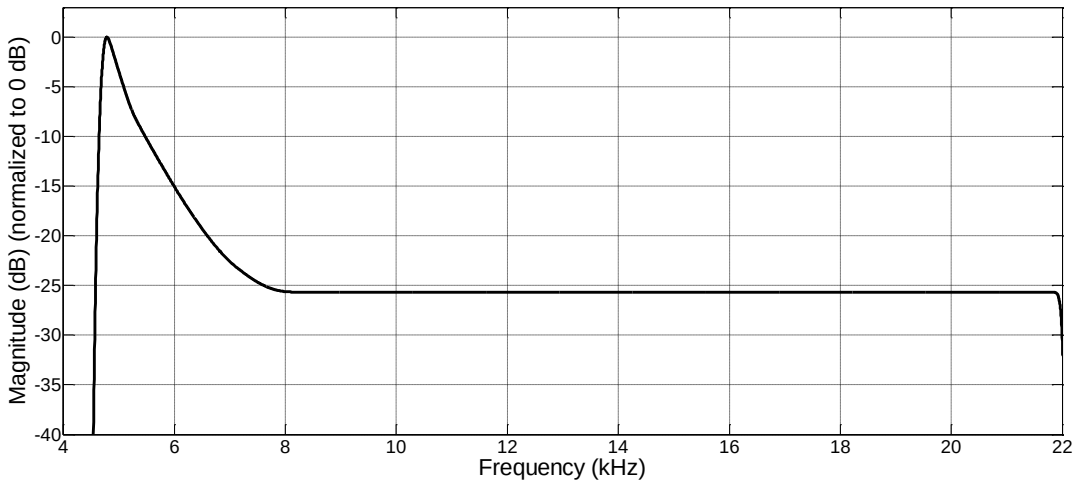


Figure 73 – Normalised magnitude response of the FIR filter constructed based on the shape of the lowest masking threshold estimated from the masker music. Its response has been flattened above 8 kHz.

Figure 74 reports the magnitudes response of the MLS prior the filtering (MLSclean); of the MLS, filtered using the aforementioned filter (MLSfilt); of the low pass filtered masker music (MUSICfilt).

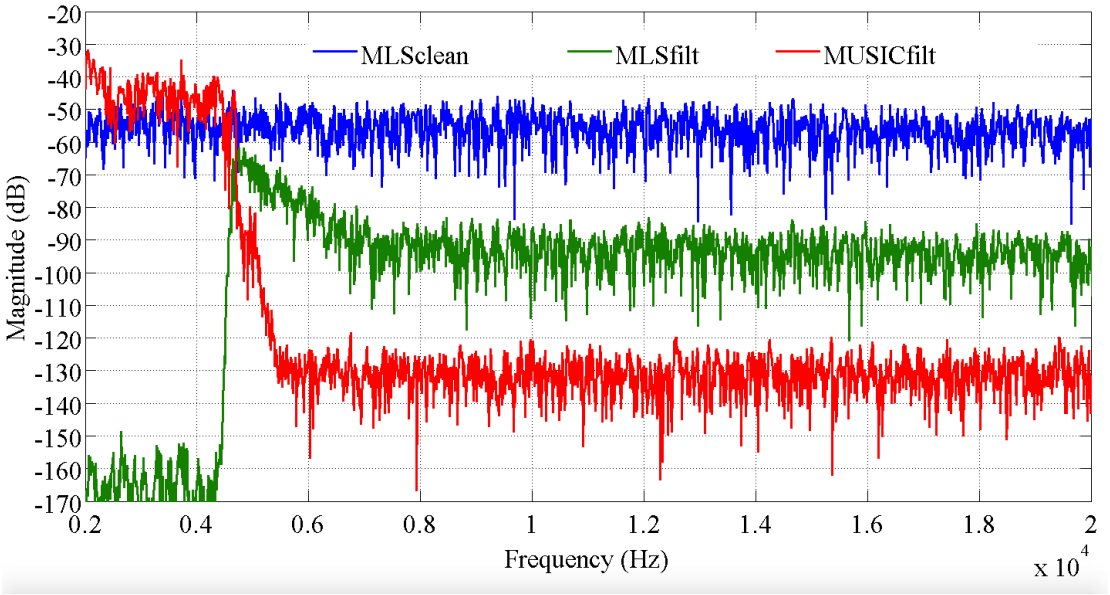


Figure 74 - magnitudes response of the unfiltered MLS (in blue); of the filtered MLS (in green); of the masker music (in red).

To avoid coloration from filtering the MLS, an "inverse" filter is also constructed to be used in the measurement process. The inverse filter is calculated starting from the former filter (hereafter dubbed forward filter) using the inversion process with regularisation described in Kirkeby & Nelson (1999). Figure 75 points out how the convolution of the two filters results in a flat frequency response (red line). Note that the inverse filter has been limited at 20 kHz, for practical reasons.

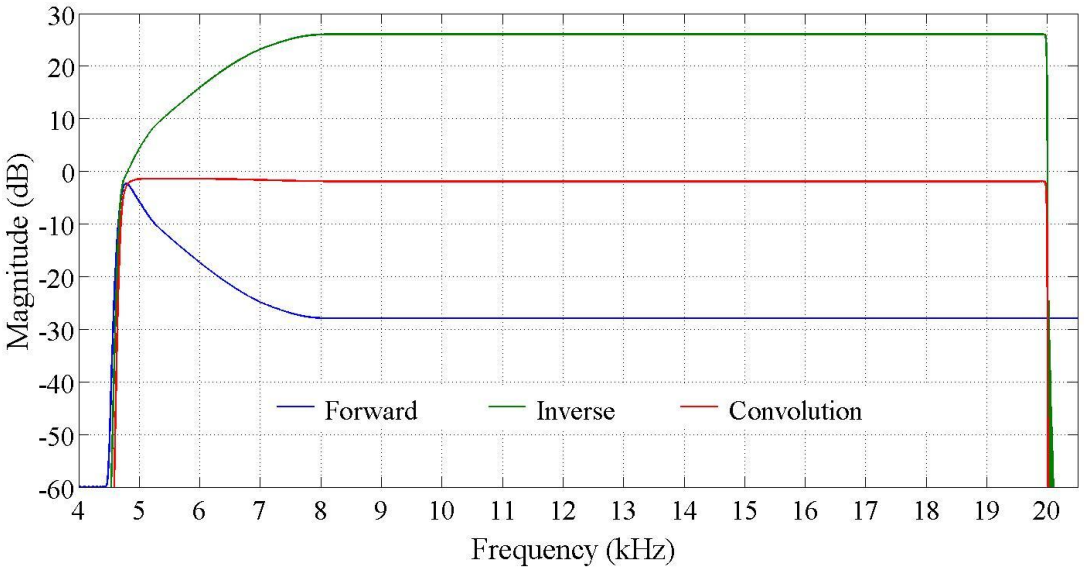


Figure 75 – Magnitude response of the forward filter (blue line); of the inverse filter (green line); and of their convoluted frequency response (red line).

5.2.4 The Masker Music

Frequency limiting the music at 4.7 kHz can retain most of the fundamental frequencies of musical instruments and most of the frequencies of the human voice. Nonetheless, the filtering will have the inevitable counter-effect of degrading how it sounds, as it will lack high order harmonics. This degradation should be taken into account when maskers have to be chosen, in order to find music that would result less affected by the filtering. However, the masker music is only intended to be a distracter

for the audience during the measurements, and it will not last longer than a few minutes. Beside its sound quality, it is important that the masker music will be effective in masking the MLS for the whole duration of a measurement. It should be considered that the more effective a masker the higher the level of the MLS that might be used. This consequently leads to a shorter measurements time or similarly to a higher SNR. To improve the masking it could be useful, for example, rising the gain at the mid-high frequencies, to increase the spread of masking. Similarly, dynamic compression of the frequencies near the cut-off point could favourite the uniformity of masking. Both operations have been applied to the masker music used in the validation chapter 6, by using compressor and equalisation audio units.

Besides these precautions, what mostly influences masking is the level of the music (although, as described by Gockel et al., a higher level does not completely results to improved masker efficiency). For example, by using very loud music (live music may commonly be played at levels of 100 / 110 dB(A) SPL in concert hall/auditoria) it should be easy to achieve a high masking threshold. However, at the same time it could be endangering, as well as annoying, for an audience to be exposed to such sound pressure levels. Moreover, such high levels are hardly obtained in listening environments like cinema, theatres and living rooms, which represent some locations where occupied impulse response measurements could be interestingly carried out. For these reasons, to prove the effectiveness of the method, measurements have been performed using a continuous A-weighted equivalent level (L_{Aeq}) of 75 dB (calculated over the duration of the music). Moreover, since the acoustic response of an environment will affect the spectral characteristic of the music, before being analysed by the psychoacoustic model, it is convolved with the impulse response of the room.

Since masking is level depended, the psychoacoustic analysis of the masker music should be representative of the actual level of the music during the measurements. To correlate the in situ level of the music with its digital representation, the A-weighted equivalent continuous level (L_{Aeq}), is used. The L_{Aeq} is defined as the total amount of energy (squared pressure) averaged in a specific time interval. It is mathematically expressed as

$$L_{Aeq} = 10 \cdot log_{10}\left(\frac{1}{t_2 - t_1} \cdot \int_{t_1}^{t_2} \frac{P_A^2}{P_0^2} dt\right) \quad [dB] \tag{65}$$

where t_2 and t_1 are the final and start measurement times (in seconds), respectively; P_A and $P_0(= 20\mu Pa)$ are the measured A-weighted and the reference pressures (in Pascal), respectively.

The correlation between the digital waveform and the reproduced music, is done by assigning a level of 75 dB L_{Aeq} to the recorded music, by relating its RMS power expressed in dB, to $P_0(= 20\mu Pa)$; then the same level is given to the reproduced music, by manually adjusting its playback gain, following the L_{Aeq} readings from a sound level meter.

5.3 Feasibility of Masked-MLS Measurements

Because of the time variance susceptibility, overly prolonged measurements are not recommended when using MLS. The duration of a measurement is equal to the length of the used sequence by the times of its repetitions. As discussed in Section 2.4.4.1, MLS measurements performed in time invariant conditions, and with uncorrelated background noise, result in an SNR increase of 3 dB for each doubling of the number of averages. Such rules can be used to forecast the time needed for a measurement to achieve a determined SNR; therefore speculating about its feasibility and accuracy.

5.3.1 Targeted Signal to Noise Ratio

Room acoustic parameters are commonly extrapolated from a noise-free EDC of at least 50 dB (thus 15 more than the range 5 dB – 35 dB used to estimate the RT30). For auralisation purposes, a free noise decay range up to 90 dB is generally recommended. Unfortunately, as documented by Stan et al., (2002), using MLS is unlikely to result in such a high SNR, due to distortion artefacts or/and time variance. Consequently, for the masked-MLS method a target of 50 dB of SNR is chosen as a more "feasible" result to be achieved in RIR measurements.

To estimate measurements time, some hypothesis has to be formulated. Vorlander & Kob (1997) pointed out that cross-correlation repacks all the energy spread in an MLS period into an impulse response, where energy is mainly at its initial part. As they stated, this gives an initial boost to the SNR proportional to the length of the used MLS and the RT of the tested venue. They also presented a formula that accounting

for the aforementioned facts, which can be used to estimate the SNR increments as a function of the number of performed averages, that is

$$\Delta snr = 10 \cdot \log_{10}\left(13.8 \cdot \frac{(N-1) \cdot T_{MLS}}{RT}\right) \quad [dB] \tag{66}$$

where T_{MLS} is the duration of one cycle of the MLS, RT is the reverberation time, and the term $(N-1)$ refers to the number of averages (the term "–1" accounts for the discarded first ML sequence). Figure 76 shows the theoretical logarithmic increment of the SNR versus the number of averages, calculated using Eq. (66).

This graph/relation can be used to determine the number of averages needed to achieve a wanted SNR, and then calculate the overall measurements times by multiplying the duration of the used MLS for the number of averages (plus one – thus the pre cycle). However, the validity of such forecast depends on the measurement scenarios, since, as mentioned, time variance and correlation of the background noise can influence the 3 dB increment rule. It has to be noted that at least for the latter condition, since in the masked-MLS method music is filtered out from the measured spectra, there should be no risk of correlated background noise within the averages.

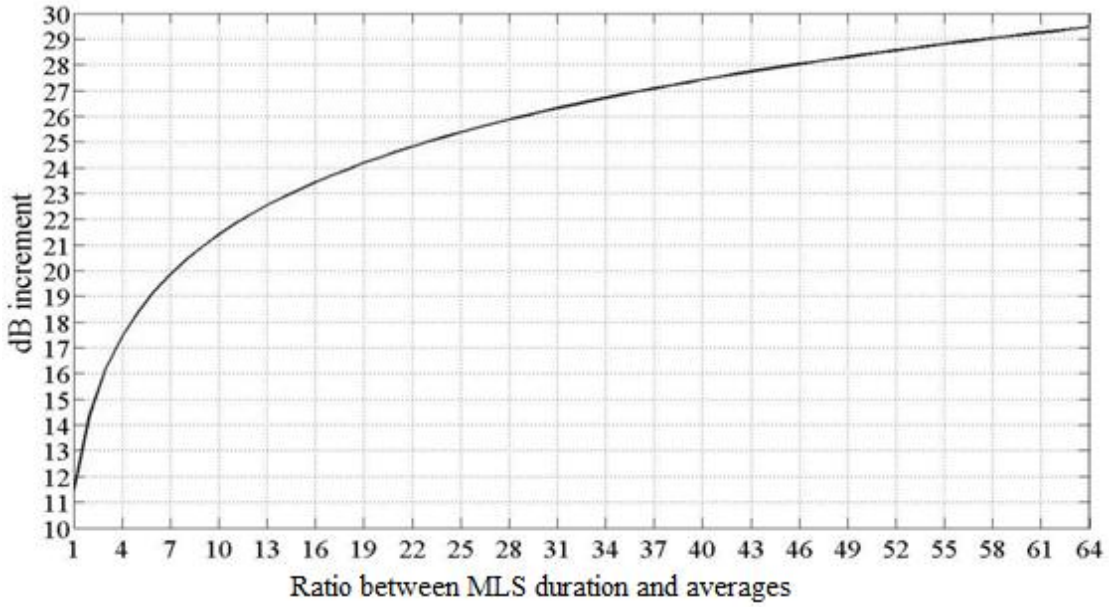


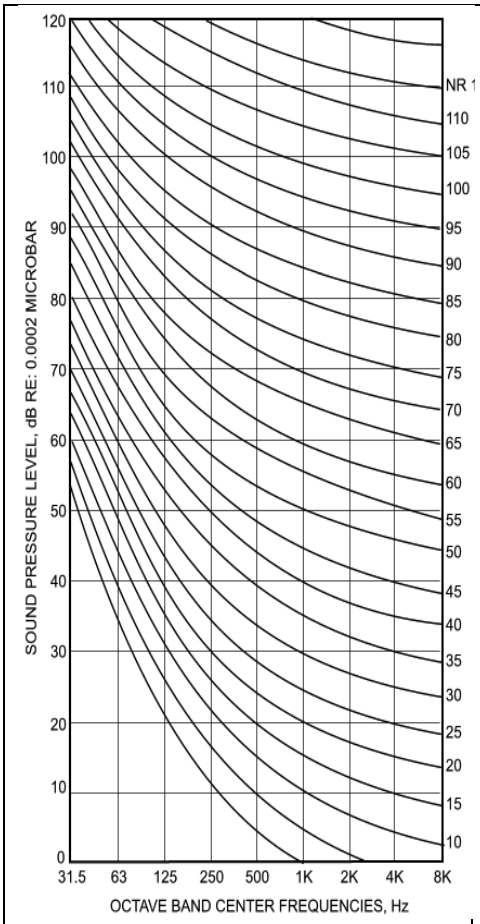
Figure 76 – SNR increment versus the number of averages.

5.3.2 Background Noise and Noise Ratio Curves

The masked-MLS method benefits from a low SNR for the range of frequency it considers, because the frequency distribution of noise in auditoria is pink or brown. Such conditions, helped on the decision of developing an MLS-based measurement method, focused on the high frequencies of the audio spectrum (a high initial SNR implies quicker measurements, which cope better with time invariant environments).

To determine the amount of averages needed to reach the target SNR, the background noise needs to be considered. To know precisely the spectra of the background noise it should be calculated in situ before measurements will be carried out. However, hypothesis could be made for venues, like theatres, cinemas, etc., since they have presumably been built in conformity to specific background noise values. To classify environments according to an acceptable indoor noise that preserves speech listening and communication without triggering annoyance, a rating system has developed,

namely the noise rating curves (NR) - detailed in the BS EN ISO 8233:2014. Such rating is mostly used in Europe, whereas another rating, namely the preferred noise criterion curves (PNC) - published in Vér & Beranek (2006), is mostly used in the USA. Architects and constructors should ensure that the background noise of built spaces conforms to the NR values reported in Table 11.



NR25	Concert halls, broadcasting and recording studios, churches
NR 30	Private dwellings, hospitals, theatres, cinemas, conference rooms
NR 35	Libraries, museums, court rooms, schools, hospitals operating theatres and wards, flats, hotels, executive offices
NR 40	Halls, corridors, cloakrooms, restaurants, night clubs, offices, shops
NR 45	Department stores, supermarkets, canteens, general offices
NR 50	Typing pools, offices with business machines
NR 60	Light engineering works
NR 70	Foundries, heavy engineering works

Table 11 – Noise rating curves and their application (extracted from http://www.engineeringtoolbox.com/nr-noise-rating-d_60.html).

A NR grade is assigned to a space based on the SPL of the highest noise level measured among the octave bands in the range (63 Hz – 8 kHz). Practically, to an SPL, a specific octave band corresponds to an NR grade as defined in the "NR table" published in the ISO 8233:2014. The highest grade among all the sub bands,

determine the NR grade of the venue. Since, NR curves are only defined up to an 8 kHz octave band, an extrapolation has been made to extend them up to 16 kHz, and allow some hypothesis to be done for masked-MLS measurements. Note that the curves are represented using the bark scale. Threshold of hearing is also reported (dashed line), and the vertical line indicates the 4.7 kHz cut off frequency point.

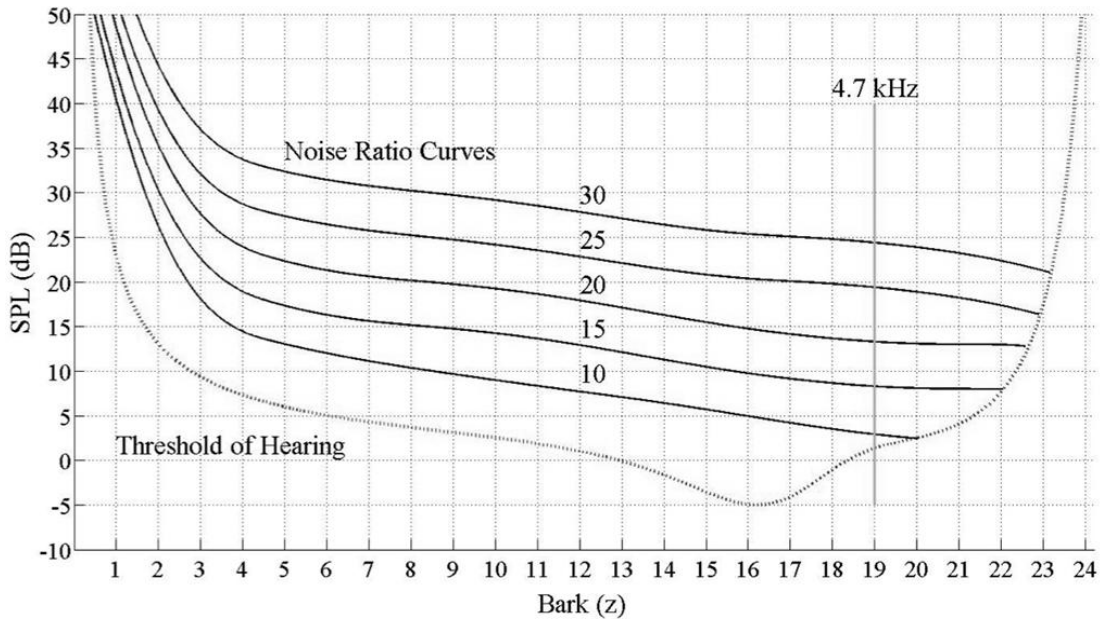


Figure 77 – Extrapolation of the noise rating curves up to the limit of the threshold of hearing, presented on a barks scale.

In the case of auditoria and concert halls, a noise SPL distribution curve rated 25 is suggested. Figure 78 shows a magnification from the 19 barks upwards. Being the level of NR25 below 20 dB SPL, it can be thought as the maximum (technical) background noise admitted for (properly built) environments, which the masked-MLS method would have to cope with in the typology of rooms specified in Table 11.

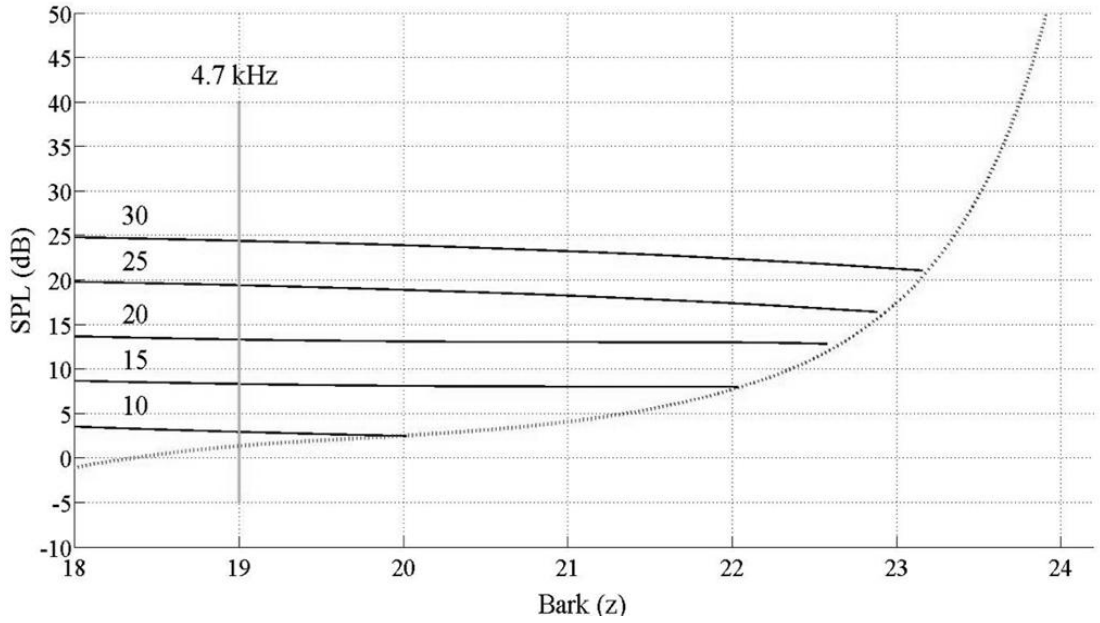


Figure 78 – Zoom on barks 18 – 24 from Figure 77.

5.3.3 Audience Noise

The overall background noise is mainly composed of an "internal" noise, which is due to all electronic equipment like HAVC, but also to some "external" noise that can be added to the background noise. The latter is generally due to construction works, roads, railways that are nearby the (not completely insulated) venue. On the other hand, the noise produced by the audience should also be considered, although, as audience is supposed to remain silent while the music is playing it generally should not contribute to increasing the background noise. However, for the sake of argument, an analysis of audience noise carried out by Jeong et al. (2012), revealed that audience noise frequency distribution reflects one of background noise. This can also be related to the Lombard effect, which is a communicative phenomenon discovered both in humans and animals for which the voice is raised or frequency is shifted/shaped in a noisy environment to raise the SNR (Pick, 1989).

Jeong concluded that audience noise, which is also dependent on the reverberation time, is generally higher than background noise in the range 500 Hz - 4 kHz, but not in large halls where audience noise is undistinguishable from background noise. Moreover, for the halls he has tested, the noise level (background plus audience) never exceeded 20 dB SPL and it dropped to 15 dB in the 8 kHz band. Hence, it can be assumed that noise distribution in the environments rated NR 25 is inclusive of audience noise.

5.3.4 Forecasting the Duration of Masked-MLS Measurements

To forecast the overall duration of a measurement needed to achieve the target SNR of 50 dB, two assumptions are made. Firstly, that the tested environment is graded NR25. Secondly, that the masker music has the masking threshold shown in Figure 79. Then, it is possible to drawn a graph relating to the MLS and the expected background, For example, Figure 79 shows the filtered MLS spectrum, the lowest masking threshold, the average SPL of the masker music, and the expected NR25 curve.

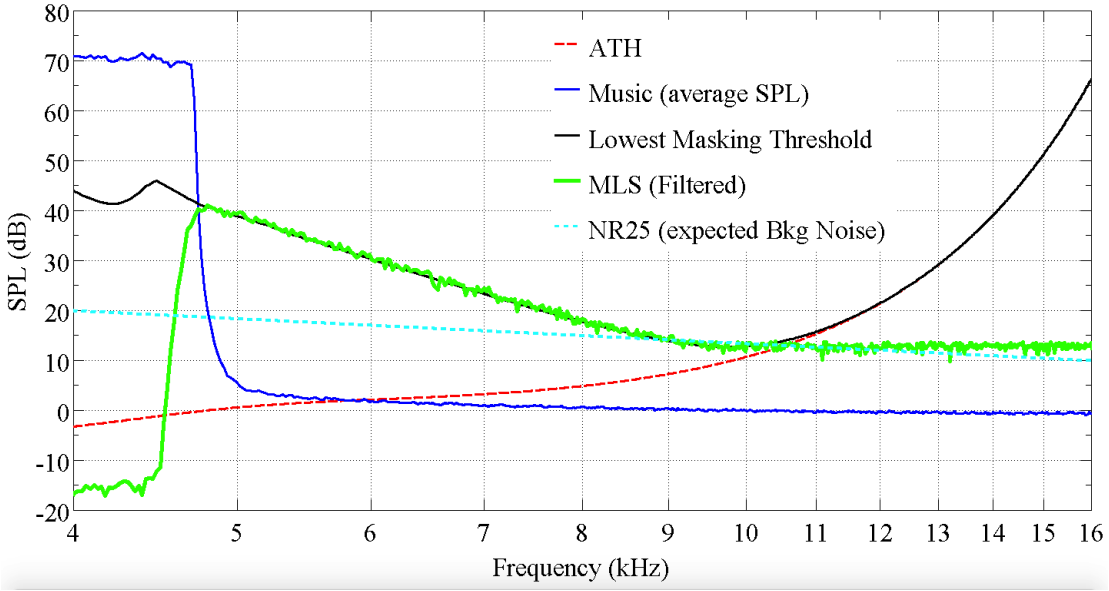


Figure 79 – Prediction of the initial SNR in a masked-MLS measurements. The blue line indicates the average SPL of the masker music the black line the lowest masking threshold estimated from the masker music; the green line the SPL of the filtered MLS; the dashed light blue line the supposed NR25-based curve; the dashed red line the ATH.

As it can be seen, the SNR is of about 20 dB around 5 kHz, and of about 0 dB at 10 kHz. Note that the MLS has a roll-off of 25 dB/octave (up to about 10 kHz). Using Eq. (66), it results that for such masked-MLS measurements, the target 50 dB SNR is reached at different times for different frequencies. For example, using an MLS sequence of 17th order (about 3 seconds at a sampling rate of 44.1 kHz), and for an RT 0.5 seconds, an increase of 30 dB should be achieved after (N=)12 averages at 5 kHz (where the initial SNR is of 20 dB). This calculation is reported in Eq. (67).

$$\Delta snr = 10 \cdot \log_{10}\left(13.8 \cdot \frac{(12-1)\cdot 3}{0.5}\right) \approx 30 \quad [\text{dB}] \tag{67}$$

Conversely, at 10 kHz, supposing an RT of 0.4 seconds, about 1000 averages would be needed to reach the dB target. Obviously, such forecast implies that the masked-

MLS method cannot successfully be used to achieve accurate measurements at a high frequency band.

The in situ measurements presented in chapter 6, will confirm this theoretical prediction, nonetheless the SNR in the tested venue was lower than the one hypothesised. Reasons are given and possible solutions, for future work proposed.

6 Validations of the Masked-MLS Method

6.1 The Measurement Scheme

Acoustic measurements software does not (generally) include the possibility of arbitrarily shaping the MLS used as probe. This limitation in controlling the pre and de-emphasis of the MLS to shape the MLS according to the estimated masking threshold, has been solved using a digital audio workstation (DAW), Audition 3.0. Such software allowed using a real time convolution module/plug-in, with custom impulse responses. On the other hand, the software ARTA (http://www.artalabs.hr - Dr. Ing. Ivo Mateljan) was used to generate and deconvolve the MLS probe. The two software programs were installed in a consumer pc (Arqivia, Intel 5th generation, 2.8 GHz, Windows 10). A Steinberg UR22 has been used as audio interface (via USB).

The software exchange audio using an internal audio routing, namely VB. Specifically, ARTA inputs an MLS probe (broadband white spectrum) to AUDITION through one channel of the virtual cable. In AUDITION, a real time convolution of the MLS with the impulse responses of the forward filter (pre-loaded in the effect plug-in) is performed. The so filtered MLS is (real time) mixed down with the masker music and output to the loudspeaker through the USB audio interface. The acoustic response of the DUT is acquired by the microphone connected to the USB audio interface, and input to AUDITION. Again in AUDITION a real-time convolution of the acquired audio with the impulse response (pre-loaded in the effect plug-in) of the inverse filter is performed, which is then directly routed to ARTA, through the second

channel of the virtual cable. Figure 80 reports a schematic representation of the discussed measurement chain.

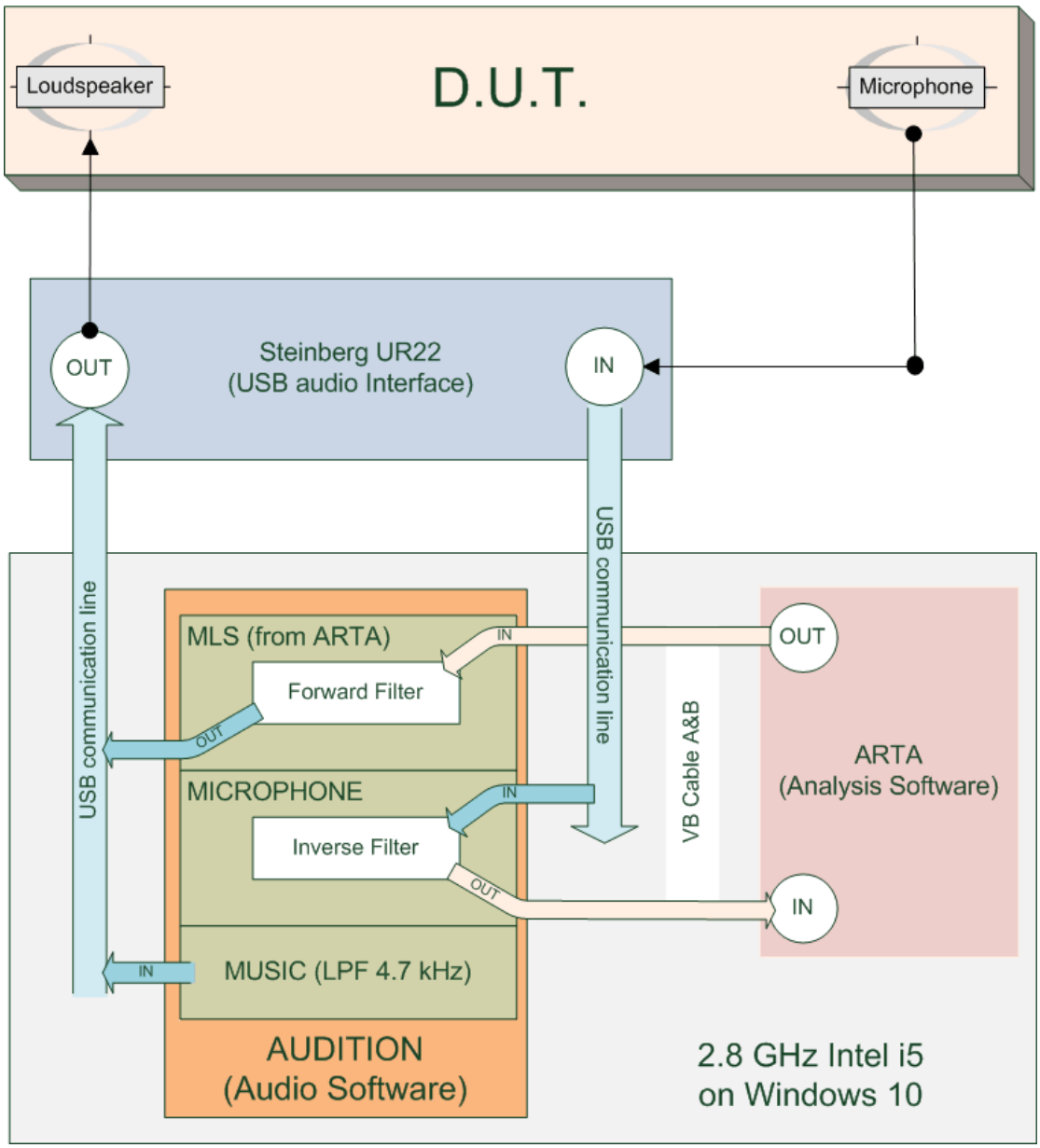


Figure 80 – Schematic representation of the measurement scheme developed to perform masked-MLS measurements. The three green blocks overlaid on the AUDITION orange block represents three (real-time) channels in the DAW. The software ARTA is used to generate the MLS probe, and to analyse the acoustic response of the D.U.T.

Figure 81 reports two screenshots of the software AUDITON, where the upper panel shows the convolution reverb module, and in the bottom panel the audio routing used to acquire and send the audio from and to ARTA. It is worth noting that, since the inverse filter is (purposely) band limited in the frequency range 4.7 kHz – 20 kHz, this has the consequential effect of filtering out the masker music (as it is low pass filtered at 4.7 kHz) from the audio acquired by the microphone. This consequently impedes that the masker music could overload (digital clipping) the (virtual) audio input of the software ARTA. This is a crucial point to allow the measurement software to be calibrated based on the actual (low) level of the received MLS, and not of the (high) level of the music.

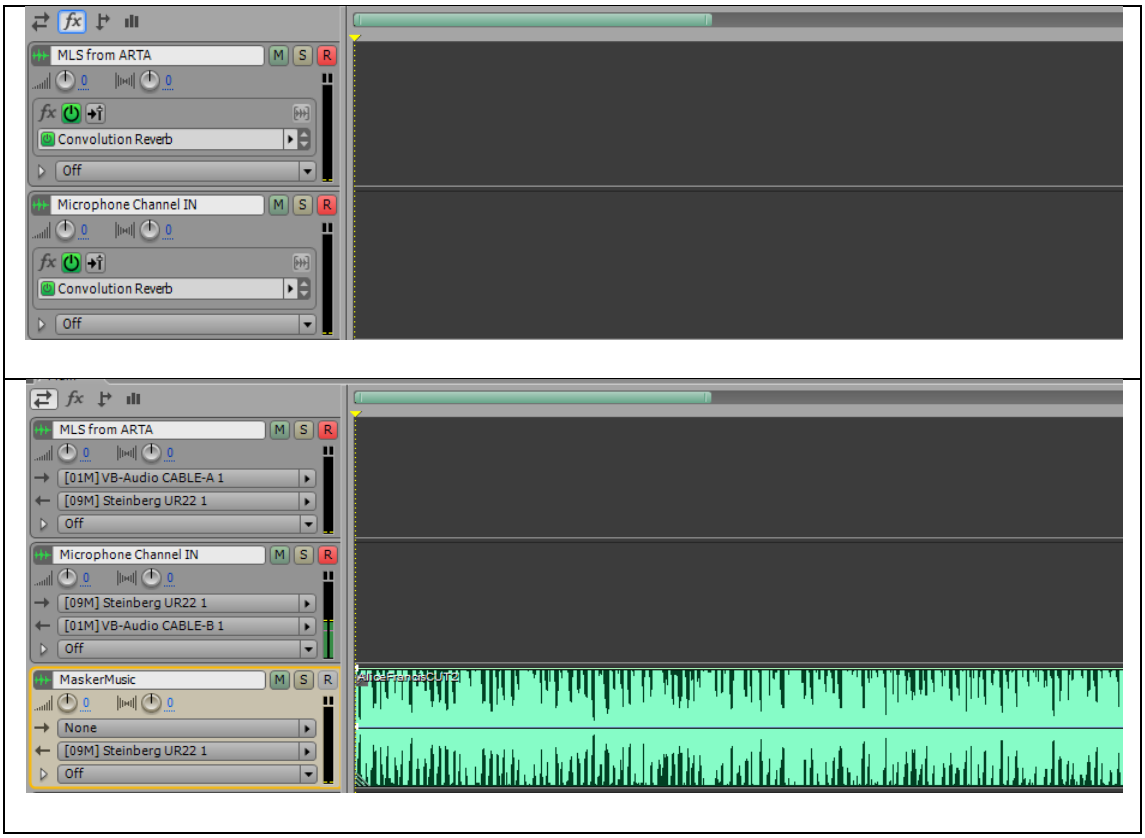


Figure 81- Screenshot of the software AUDITION where in the (upper panel) the FX module used to perform the convolution with the forward and inverse filter is represented. In the bottom panel shows how the audio routing is performed.

Although, the forward and inverse filter have been constructed has a linear phase FIR filters (thus only a delay is introduced). The effectiveness of the proposed measurement scheme has been proven to secure that unwanted non-linearity could invalidate the measurements. To do so, the frequency responses of a loudspeaker were compared with, or without, using the filters. Figure 82 shows overlaid the frequency responses of the loudspeaker for three different MLS measurements. The first, used as reference, was carried out using an MLS without any filtering (blue line), the second by using only the forward filtering (orange line), and the third by using both the forward and the inverse filtering (red line). As it can be seen, the latter measured frequency response matches the reference measurement in the range 4.7 kHz – 16 kHz. To feed accurate comparisons of the two frequency responses, Figure 83 reports their coherence and SNR analyses. Note that since filtering changes the gain of the signals, the level of the two channels of the DAW (MLS and MIC) have been manually adjusted to achieve the same level of the reference measurements. Such calibration has been repeated before any measurements sessions.

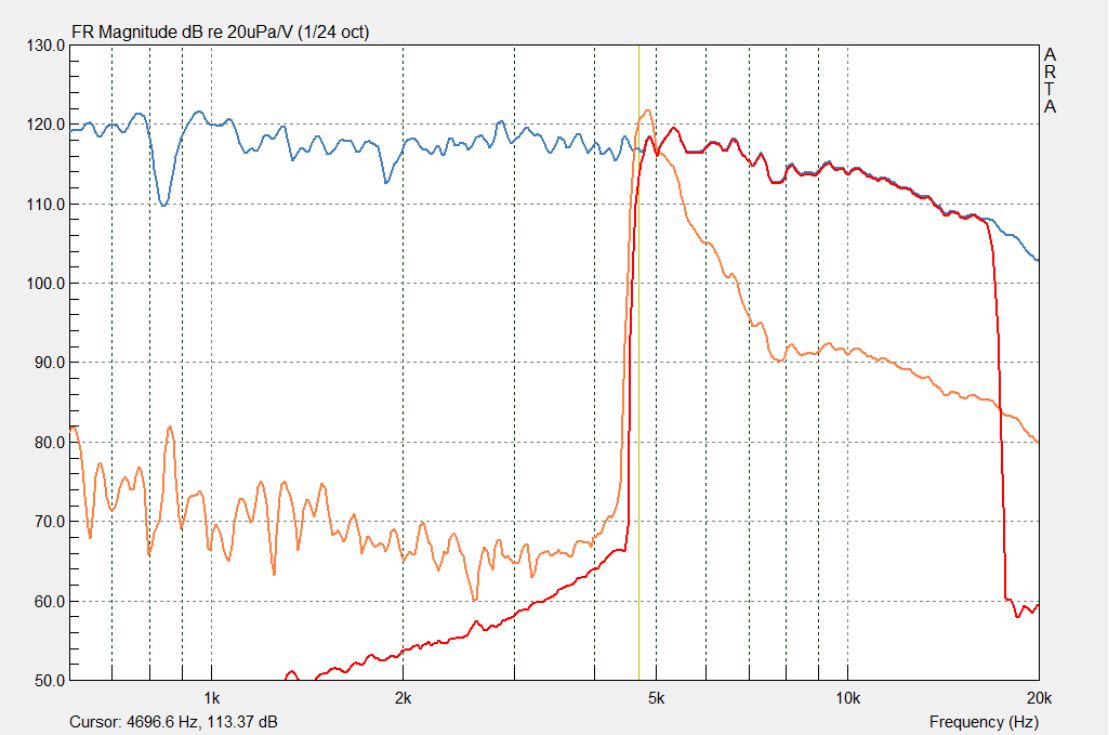


Figure 82 – Overlaid frequency responses of a loudspeaker measured using three different MLS configurations. The frequency response in blue has been calculate using a broadband MLS; the one in orange, using only the forward filter on the MLS; the red measurements by using both the forward and the inverse filter on the MLS.

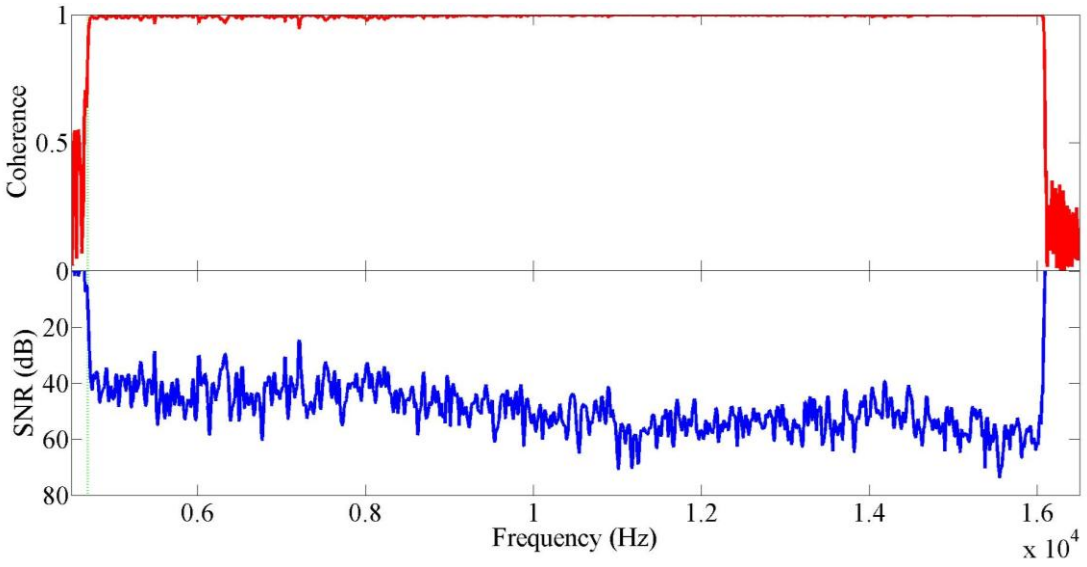


Figure 83 – Coherence and SNR analyses of the reference FRF, and the one measured using the double filtering on the MLS probe.

6.2 Used Masker Music

The (pop/dance) song “shoot him down!” from the swing singer Alice Francis has been chosen as the masker music. This genre of music commonly presents a low dynamic range and few low-volume/silent parts. The presence of a feminine voice also guarantees the presence of mid-high frequency harmonics that could favourite a forward masking effect. Nonetheless, low pass filtering the music decreases its sound quality, the mid-low frequency rhythmic sounds (drums, samplers and electronic instruments), as well as much of the frequencies of the voice. The masker music has been judged “duller” by the participants at the listening tests, which could be a limitation of the masked-MLS method. However, this aspect is of a secondary importance, since is only intended to be used as a distraction while the measurements are being performed.

The song is about 3.5 minutes long, but only a 50 second long extract has been used for the validation measurement. This excerpt presents a dense frequency response and the presence of harmonics near the cut of point (4.7 kHz), as a performed spectral analysis has highlighted. The piece of music has been equalised and dynamically processed, prior to the measurements. It has then been consecutively repeated for the whole duration of the measurements, which lasted for, any presented case scenario, two minutes and a half on average. Figure 84 shows the normalised PSD of the excerpt of music.

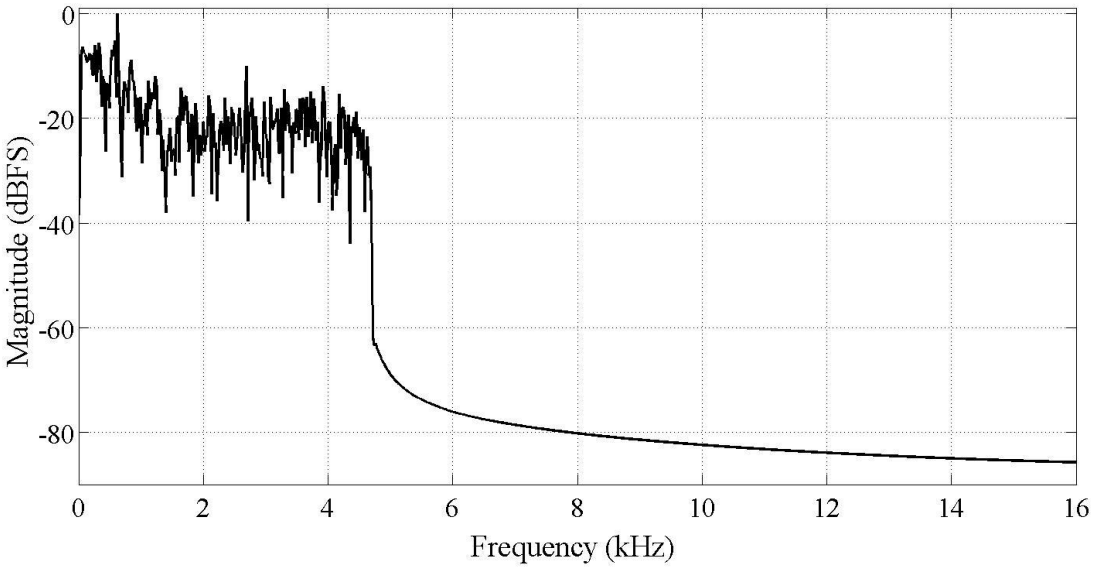


Figure 84 – Normalised PSD (using Welch periodogram) of the used masker music.

6.3 Tested Venue: Broadcasting Studio “Sala M”

Figure 85 shows the control room where the validation measurements have been performed. The technical descriptions of the loudspeaker and the microphone used for the measurements, being the same used to validate the presto-chirps method, are reported in Section 4.3. One of the two microphones, visible in the figure, was connected to the NC10 sound level meter. The other, was connect to the audio interface, and used for the actual measurements.

Figure 85 - Photograph of the room used to perform the masked-MLS measurements.

The references (band limited) RIR has been measured using a 17^{th} order 70 dB(A) MLS, averaged 12 times. Figure 86 shows, in the upper panel, the reference RIR, and in the bottom panel, its FRF.

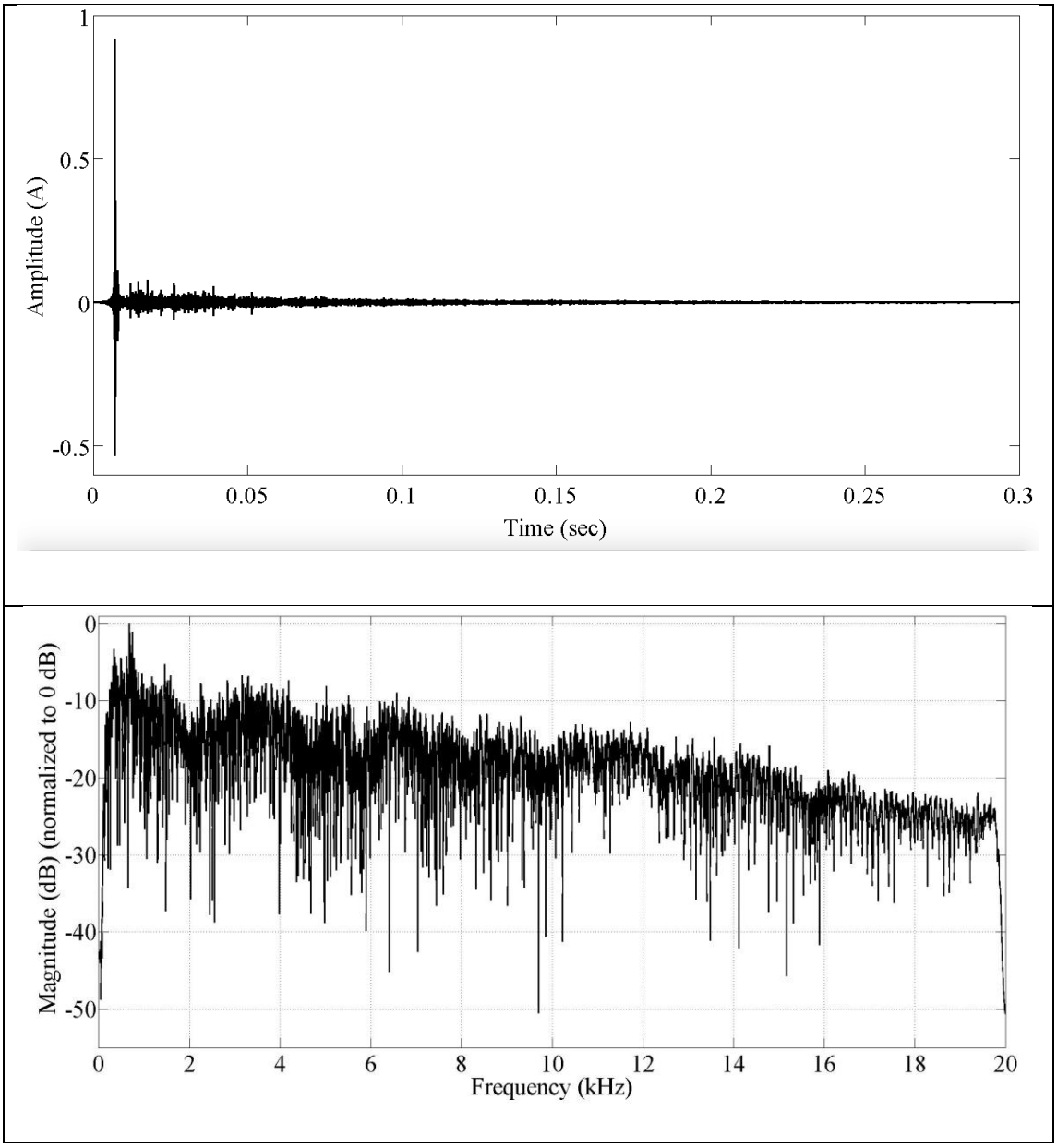


Figure 86 – Reference RIR (upper panel), and FRF (bottom panel) of the tested room.

6.3.1 Listening Tests

Figure 87 reports the lowest masking threshold resulting from the psychoacoustic analysis of the masker music used in this presented validation. It also reports the spectrum of the (forward) filtered MLS; the average SPL of the masker music; and the expected background noise distribution graded NR25, which is a typical case for broadcasting rooms.

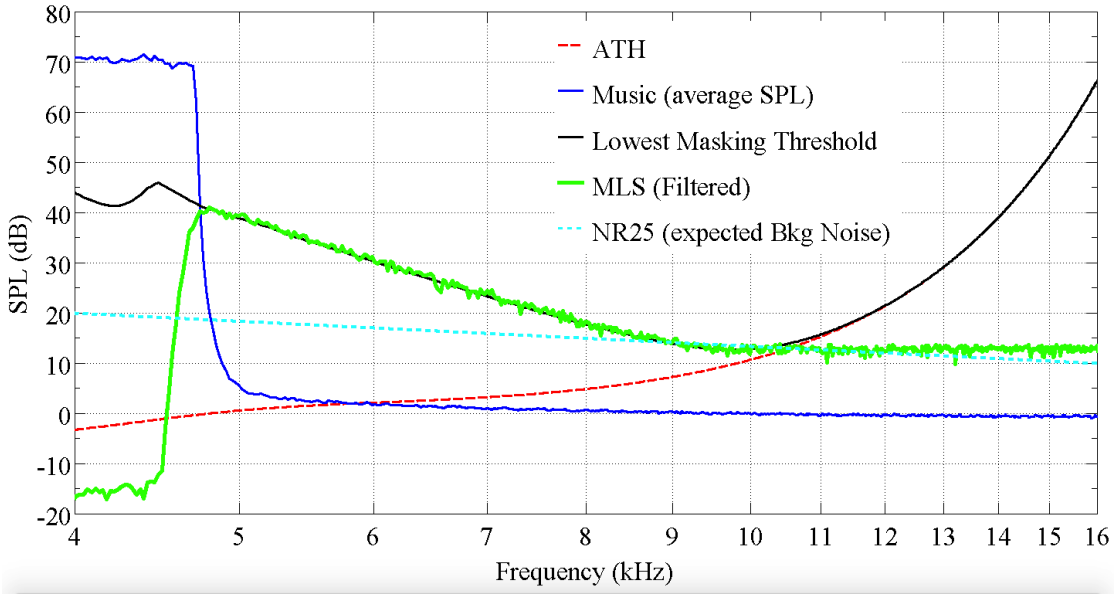


Figure 87 – Prediction of the initial SNR in a masked-MLS measurements. The blue line indicates the average SPL of the masker music the black line the lowest masking threshold estimated from the masker music; the green line the SPL of the filtered MLS; the dashed light blue line the supposed NR25-based curve; the dashed red line the ATH.

The psychoacoustic analysis of the masker music, suggests that the filtered MLS should results inaudible when played at 40 dB SPL. The MLS, only theoretically possesses a crest factor of one, but in practice it might easily exceed 10 dB. The high-level peaks of MLS, could then result audible if they exceed the masking threshold. Therefore, to firstly address this issue, the playback level of MLS has been set 10 dB

lower that the predicted masking threshold. Secondly, listening tests were carried out to evaluate if masking had actually occurred.

Listening tests have therefore carried out starting from an SPL level of 30 dB at the 5 kHz third-octave band, linearly averaged over the length of the measurement. The MLS was filtered using the forward filter, and reproduced as described in the measurement scheme shown in Section 6.1.

The listening tests were conducted on 10 experienced sound engineers (all actively working in a broadcasting company, and certified having normal hearing). The tests were carried out in the same room where the masked-MLS measurements have been performed. Participants were seated at the same position were the calibration of the music and MLS have been performed. Music was played at a level of 75 dB L_{Aeq}, while the (forward filtered) MLS was initially reproduced at a level of 30 dB at the 5 kHz third-octave band, linearly averaged over the length of the measurement.

A brief explanation of the listening test has been given to the attendees to instruct them to declare if the noise was or was not audible. After the first used MLS level resulted audible, its level was progressively reduced until inaudibility was confirmed. When this had occurred, the L_{Aeq} of the MLS was measured through a sound level meter, and results stored for a later analysis. This test was repeated twice for each participant. Therefore, a total of 20 L_{Aeq} values were obtained at the end of the session. All tests were performed on the same days, with only one participant at a time and the (same) operator present.

The lower measured L_{Aeq}, among all the collected values, has been declared the inaudibility level, for the tested venue, and with the used measurement instrumentations. It corresponded to 28 dB L_{Aeq}.

A second cycle of listening tests was run using the same participants and the same testing scenario. In this test, attendees were briefed to declare until when they subjectively felt the MLS, whose level would have been gradually raised, as "not disturbing" for the listening of the music. From the twenty collected answers, the median value was taken. It corresponded to an L_{Aeq} of 37 dB (participants described the noisy music as from an old recorded material).

Following the tests, with only the operator present on the tested venue, A-weighted octave bands SPL measurements were done using the NC10 sound level meter. In addition, $1/24^{th}$ octave bands measurements were performed using the software ARTA. These (linearly averaged) SPLs measurements were carried out when only the music or the MLS, for the two L_{Aeq} levels determined in the listening tests, were being reproduced, and the background noise. Figure 88 and Figure 89 reports the SPL analyses, in the third octave bands, and in the $1/24^{th}$ octave bands, respectively.

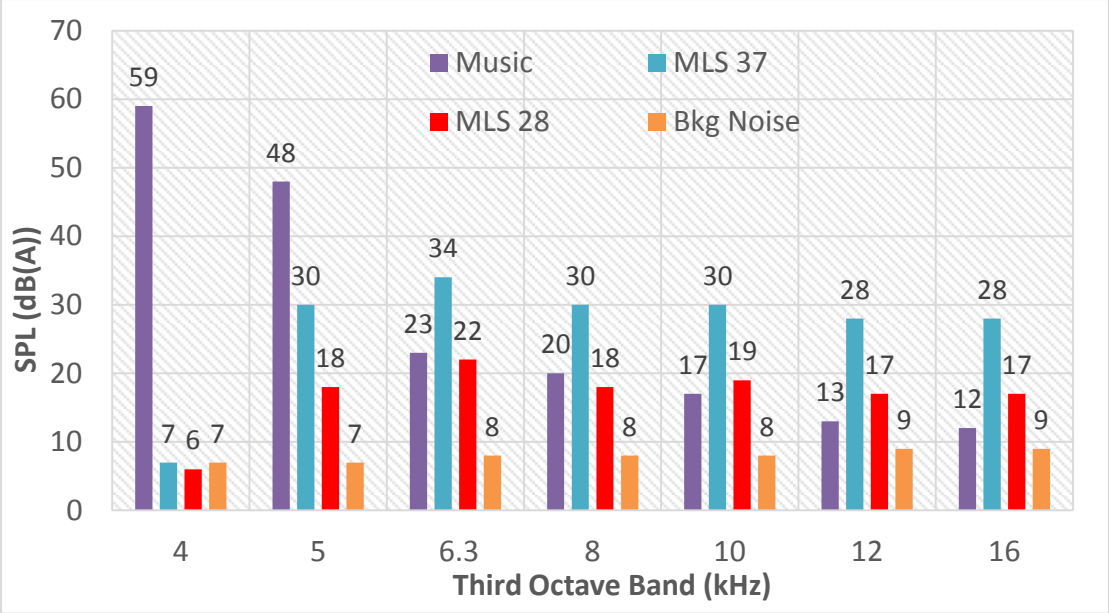


Figure 88 - A-weighted linearly averaged SPLs of the masker music, the 28 and 37 dB L_{Aeq} MLS, and the background noise. The measures were carried out using the sound analyser NC10 and refer to an 1/3rd octave band analysis.

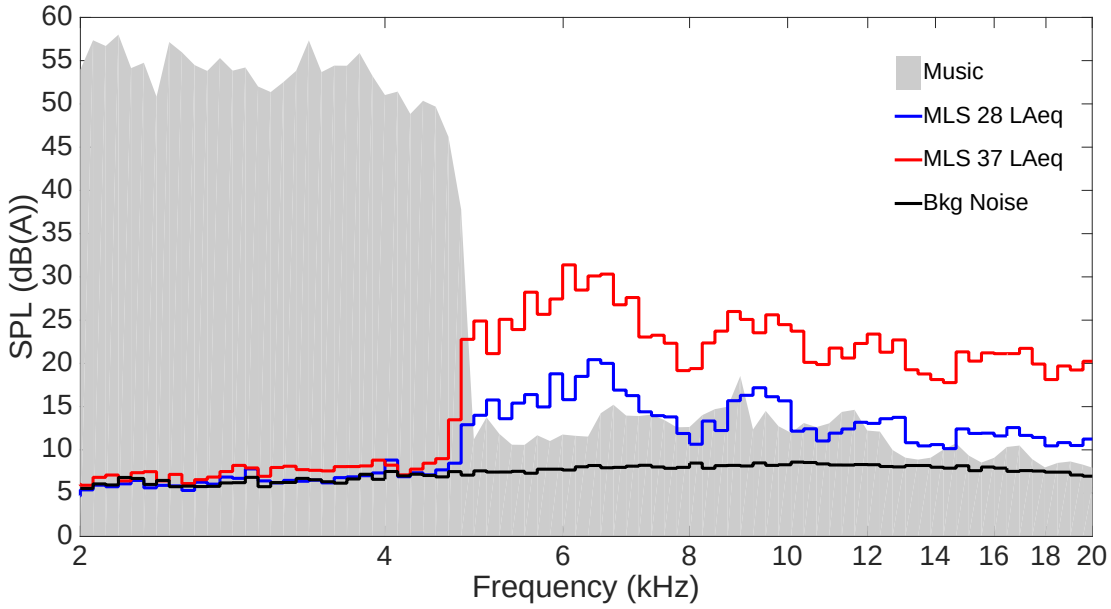


Figure 89 - A-weighted linearly averaged SPLs of the masker music, the 28 and 37 dB L_{Aeq} MLS, and the background noise. The measures were carried out using the software ARTA and refer to an 1/24th octave band analysis.

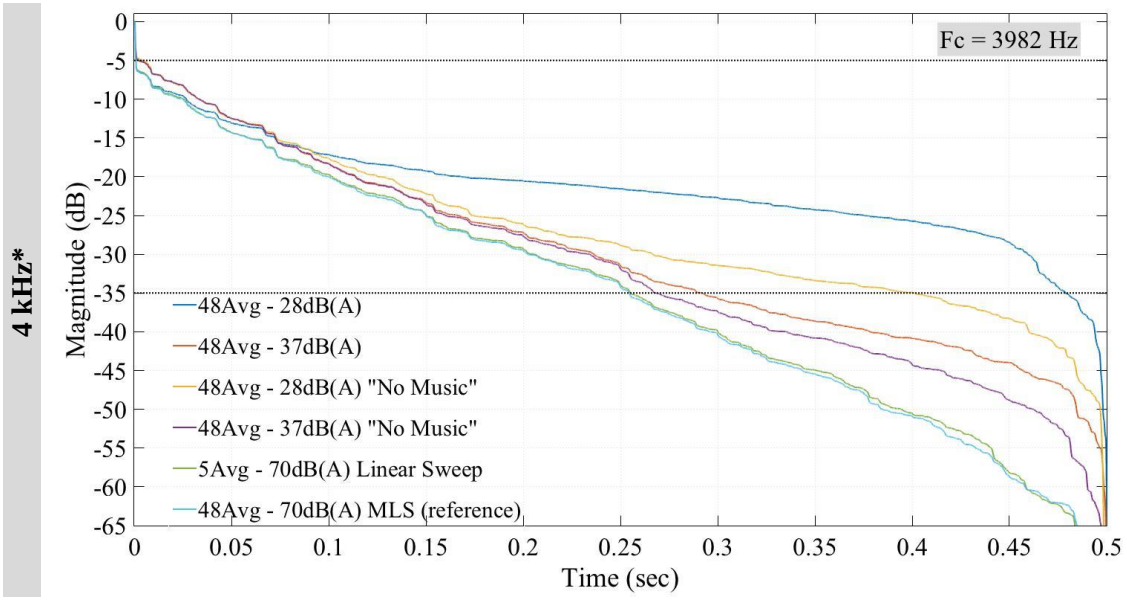
From the reported measurements, it can be seen that the spectra of music presents peaks in the frequency range above the 4.7 kHz cut off point. This is caused by the non-linear behaviour of the loudspeaker, which creates intermodulation distortion (IMD), and this result in a measurements noise higher than just the background noise. Moreover, the spectra of the MLS, being modified by the acoustic response of the room, will have low SNR where it presents local minima.

6.3.2 In situ Measurements

MLS measurements were carried out using an MLS of 17^{th} order (131071 samples), which for a sampling frequency of 44100 Hz results in about a 3 second-long sequence. The MLS was averaged 48 times. These results, including the pre-fetch cycle, were about 144 seconds-long measurements (slightly less than 2 and half minutes). Such duration should keep the risk of time variance low, which however, should be verified at the end of the measurement process. For example, Vorlander & Kob (1997), suggested comparing the autocorrelation functions from each average to check for time variance problems.

Masked-MLS measurements were carried out for both the discussed levels, thus 28 and 37 dB L_{Aeq}, with and without the masker music being played. They were then compared with two reference RIRs. The first reference RIR was measured using a 70 dB L_{Aeq} MLS, averaged 48 times. The second reference RIR was measured using a 70 dB L_{Aeq} linear sine sweep, averaged 5 times. Figure 90 shows the EDCs, in the octave bands centred at 4, 8, and 16 kHz, for all the afore mentioned measurements.

Measurements performed using the 28 dB(A) MLS (blue line) and 37 dB(A) MLS (orange line) played with music, clearly did not achieve sufficient SNR during the averaging process; for all the tested sub bands. It is interesting to note that for the measurements performed without the masker music, the 28 dB(A) MLS (yellow line) shows a sensitive reduction of the noise; on the contrary, for the 37 dB(A) MLS (purple line), the difference is relatively low. These results on the one hand demonstrate that the inverse filtering can efficiently exclude all the frequencies below 4.7 kHz from the acquired sound. On the other hand, the IMD of the loudspeaker can further reduce the SNR. Moreover, it can be deducted that for both the tested levels the masked-MLS method failed to achieve the 50 dB SNR target.



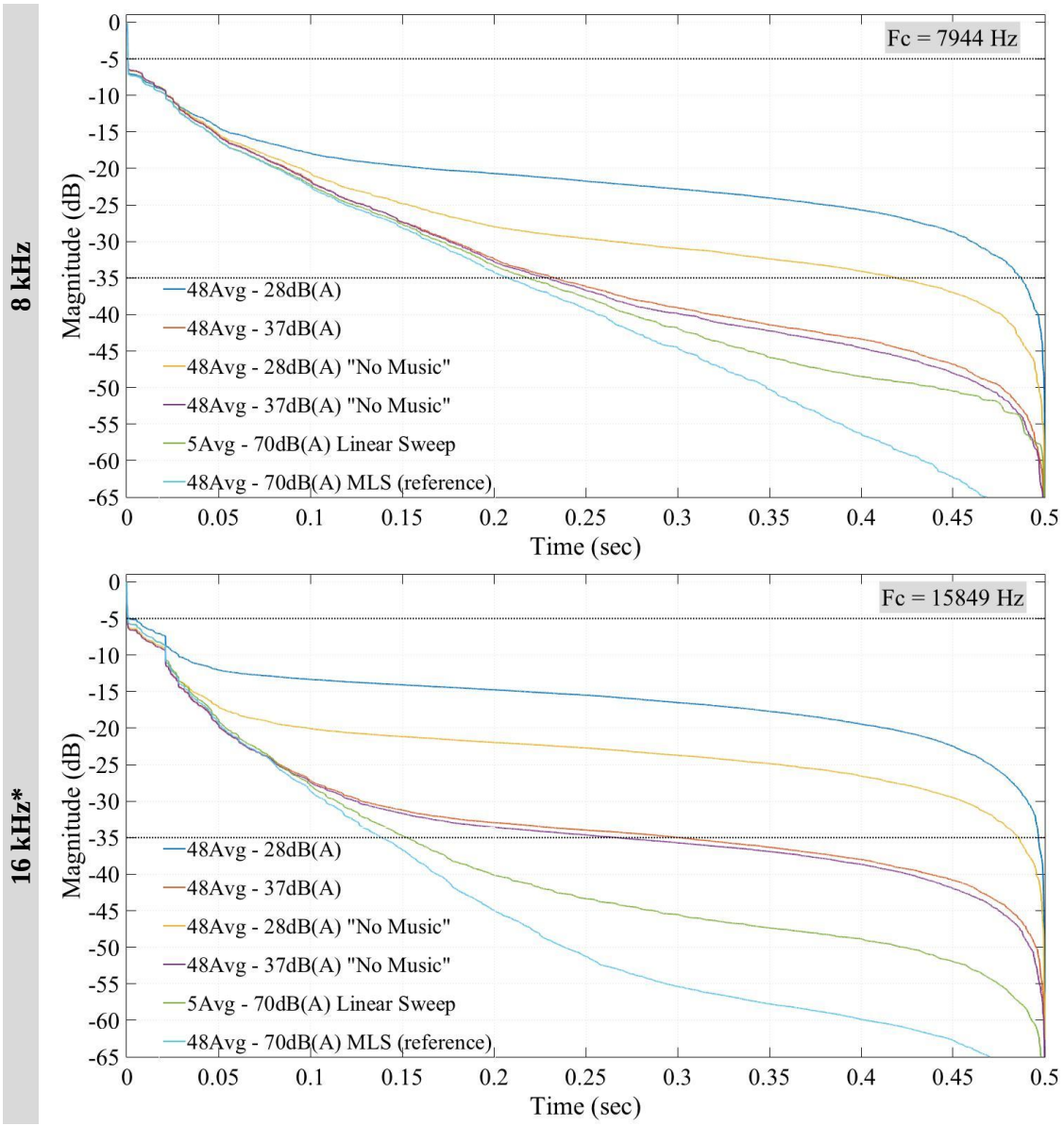


Figure 90 – EDCs curve for the measured RIRs. Each panel relate to a specific octave bands as labelled in the figure. The EDCs are related, from top to bottom, to the: RIR measured using a 28 dB L_{Aeq} masked-MLS, played with the masker music. RIR measured using a 37 dB L_{Aeq} masked-MLS played without the masker music. RIR measured using a 28 dB L_{Aeq} masked-MLS played with the masker music. RIR measured using a 37 dB L_{Aeq} masked-MLS played without the masker music. Reference RIR measure using a 70 dB L_{Aeq} 3-seconds long linear sweep, averaged 5 times. Reference RIR measure using a 70 dB L_{Aeq} 3-seconds MLS, averaged 48 times.

Table 12 reports the comparisons between the acoustical parameters EDT, RT20 and RT30, from the reference RIR measured using the 70 dB L_{Aeq} MLS, and the RIRs measured using the masked-MLS method for the L_{Aeq} levels of 27 and 38 dB. The

reference parameters are given in seconds, while for the MLS cases are given as difference with the reference parameters. The values in red are outside the DL (= 3%) of the specific parameter. In the rightmost block, the SNR of the three measurements is reported. The acoustic parameters and the SNR have been calculated using the plug-in AURORA. Results highlight that only for the MLS played at 38 dB L_{Aeq} the acoustic parameter EDT and RT20 could be estimated with an error within the DLs.

Param.	**EDT [sec]** *(DL = 0.024 s)*			**RT20 [sec]** *(DL = 0.024 s)*			**RT30 [sec]** *(DL = 0.024 s)*			**SNR [dB]**		
Bands	**Ref.**	**37dB**	**28dB**	**Ref.**	**37dB**	**28dB**	**Ref.**	**37dB**	**28dB**	**Ref.**	**37dB**	**28dB**
4k*	**0.41**	0.05	-	**0.57**	-	-	**0.61**	-	-	**54**	**13**	**3**
8K	**0.26**	0.01	0.01	**0.36**	0.01	0.10	**0.44**	0.02	0.27	**56**	**33**	**15**
16K	**0.29**	0.01	0.02	**0.21**	0.00	0.09	**0.26**	0.02	0.26	**51**	**28**	**10**

Table 12 – The columns labelled "Ref" report the acoustic parameters, calculated using the plug-in Audacity, for the reference RIR (measured using an MLS). The columns labelled 37 dB report the differences with the reference values for the RIR measured using a 28 dB L_{Aeq} masked-MLS, averaged 48 times. The columns labelled 28 dB report the differences with the reference values for the RIR measured using a 28 dB L_{Aeq} masked-MLS, averaged 48 times.

Figure 91 shows the cross-correlation between the reference RIR and the RIRs measurements using the 27 and 38 dB L_{Aeq} masked-MLS. This analysis shows that both the measured RIRs are severely affected by noise, and although the latter is slightly better than the former measured RIR, is still cannot produce an accurate measurement of the RIR. The presence of the noise in the measured RIRs, can be clearly seen in magnification of their tails shown in the yellow box within the figure.

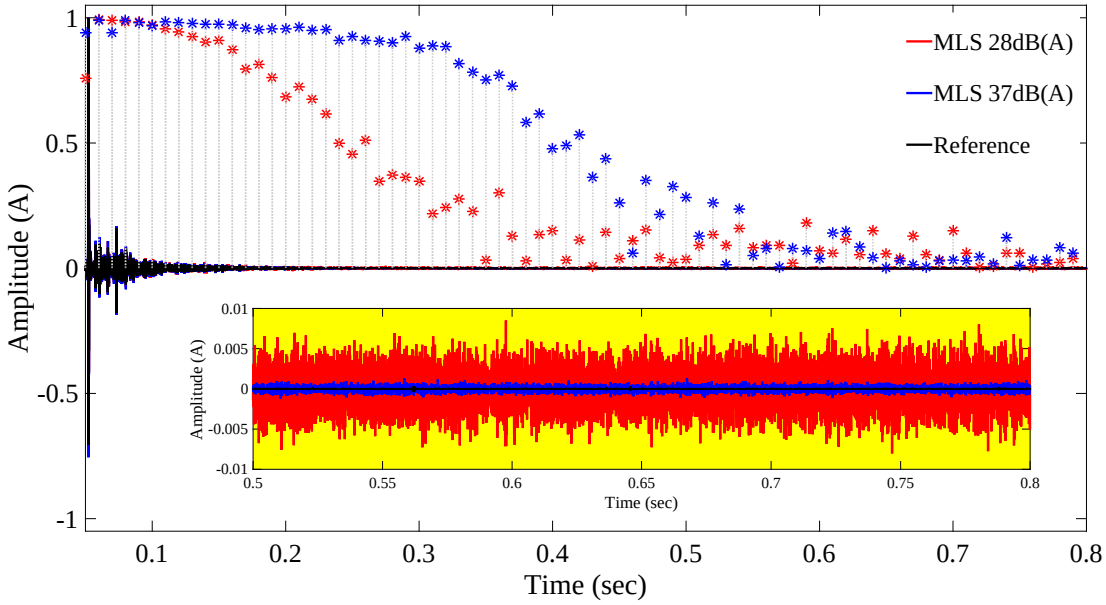


Figure 91 – Correlation analysis between the reference RIR and the RIRs measured using a 28 dB L_{Aeq} masked-MLS (blue stars); and a 37 dB L_{Aeq} masked-MLS (red stars).

Figure 92 shows the coherence analysis (in the range 5 kHz – 20 kHz) between the reference FRF and the FRFs measurements using the 27 and 38 dB L_{Aeq} masked-MLS, in the upper and bottom panels, respectively. Clearly, for the first case the noise strongly biases the measurement, in particular in the frequency range above 15 kHz, which are likely to be caused by the low energy fed by the loudspeaker at that range. For the second case, although the noise is less detrimental, the frequency response of the measured RIR is still inaccurate, especially at higher frequencies.

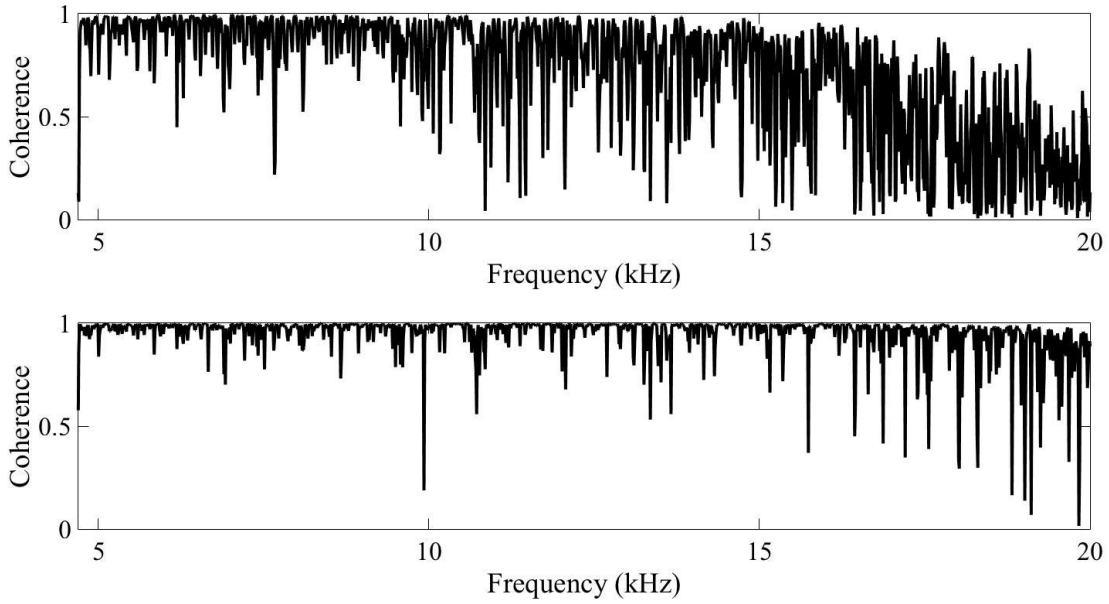


Figure 92 – Coherence analysis between the reference MLS RIR and the RIR measured using a 28 dB L_{Aeq} masked-MLS (upper panel); using a 37 dB L_{Aeq} masked-MLS (bottom panel).

6.4 Discussions

Listening tests have highlighted that to perform inaudible measurements the level of the masked-MLS probe, frequency as the masking threshold produced by a masker music, should be of 27 dB L_{Aeq}. However, measurements performed in situ have proven that the SNR achieved averaging a 17^{th} MLS 48 times, is still too low to allow estimating acoustic parameters with sufficient accuracy. A similar result was also obtained for a masked-MLS probe of 38 dB L_{Aeq}, although the SNR was clearly higher. Several reasons cause the presented method to fail. Firstly, the frequency shape given to the MLS, which is of about -25 dB/octave in the range 4.7 kHz – 10 kHz, is too excessive to permit a sufficient SNR increase after a reasonable measurement duration. Secondly, both the intermodulation distortions and the acoustic response of the room decreased the SNR.

In light of this, the presented method should be redesigned, and new measurements performed, to verify if masked-MLS measurements could be feasible under different settings. For example, the filter used to shape the MLS could include the (inverse) frequency response of the tested environment. In addition, the SNR could be increased for the frequencies above 10 kHz by using a forward filter that accounts for the masking threshold in that frequency range. However, using a different frequency shaping requests new listening tests to be done.

Although, the suggested improvements might produce better results, they do not guarantee that performing inaudible measurements would be possible.

7 The Hybrid Method

The final step of the hybrid method consists on merging the impulse responses calculated using the presto-chirps and the masked-MLS methods. Recombining RIRs measured in complementary frequency region, furthermore measured using different software, requests to overcome some difficulties.

7.1 Validating Measurements

The two methods have been used to calculate the RIR of the same room used for the masked-MLS measurements; details can therefore be found in chapter 6. For the presto-chirps method, a 60 dB L_{Aeq} musical stimulus, presented in Section 3.6.3, has been used. For the masked-MLS method, since it has been proven inaccurate for the tested levels of 28 and 37 dB L_{Aeq}, in this validation a 50 dB L_{Aeq} probe has instead been used. Two references measurements were also carried out, the first using a 17^{th} order MLS, averaged 48 times; the second using a 10-second linear ESS, averaged 5 times. Both probes had an L_{Aeq} of 70 dB, and have both been compared with the RIR measured using the Hybrid method. Figure 93 shows, in the upper panel, the reference MLS-measured RIR of the tested room; in the bottom panel, the FRF.

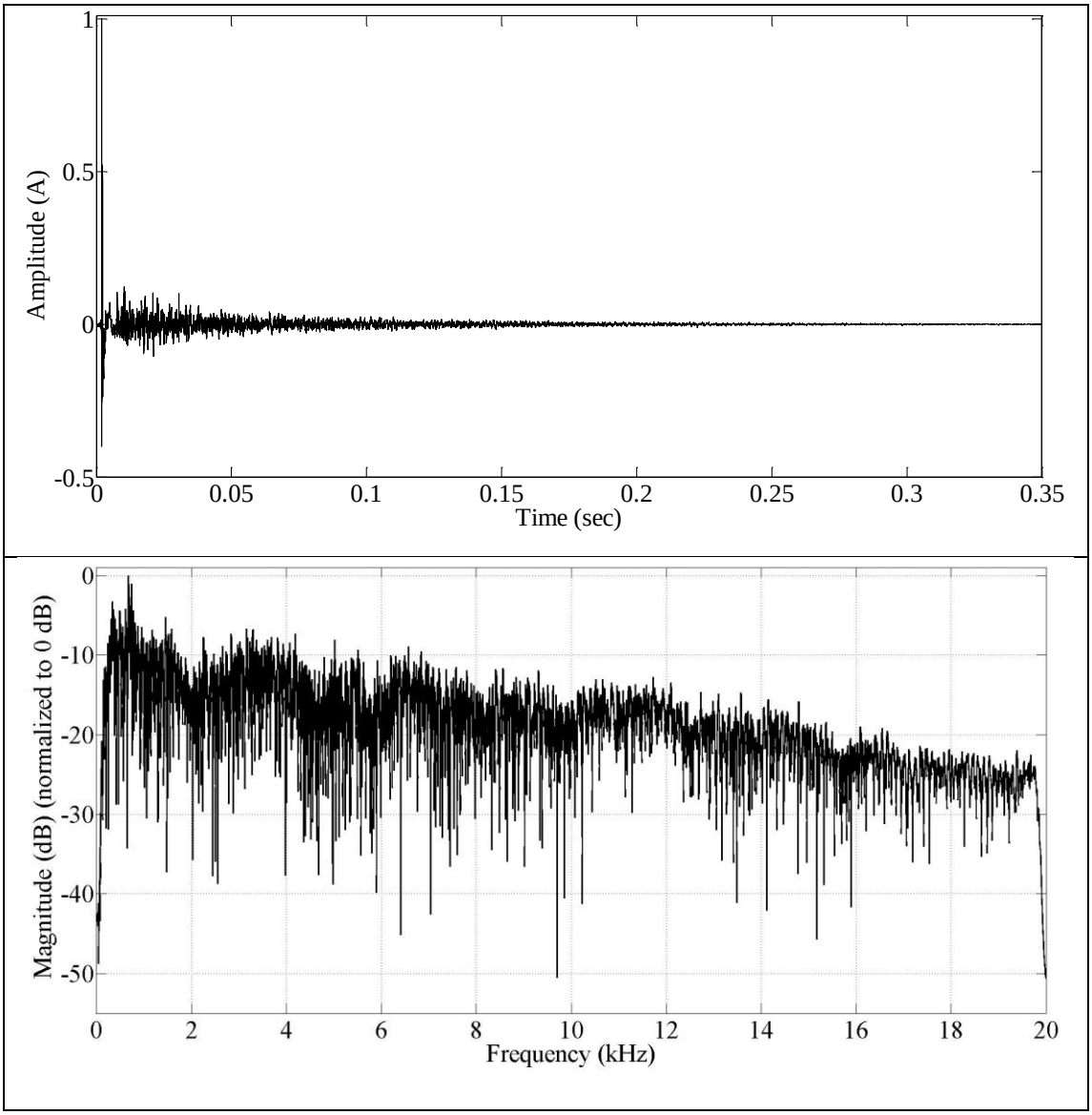


Figure 93 - Reference RIR (upper panel), and FRF (bottom panel) of the tested room.

7.2 Time Alignment of the RIRs

Misaligning the waveforms of the two band-limited RIRs might cause transient peaks or other forms of distortion to bias the recombined RIR. Therefore, it is of primary importance that the two bands limited RIRs are perfectly time aligned before being summed up.

Cross-correlation analysis could theoretically be used to determine the time lag between the two limited RIRs. However, since they only have a limited portion of the audio spectrum in common, (around 4.7 kHz), results are often unreliable. This is because they waveforms are too dissimilar. Notch filtering could be used to filter the two hybrid-measured RIRs, around the cut off point, to let the two impulse responses look similar. However, when using this approach, results were not satisfactory.

It has therefore been preferred to use a "manual" time alignment of the RIRs. This consisted on firstly filtering a broadband reference probe (in situ measured) with a two-channel filter bank, which gives two band limited RIRs with spectra equal to the ones of the hybrid-method measured RIRs. The last two are filtered too using the two-channel filter bank to ensure that their spectra are perfectly complementary. The filter bank consists of a low pass and a high pass filter channel, both constructed using an FIR filter (constrained least-squared design algorithm, 301 taps, 0.1 dB of pass band ripples, 30 dB attenuation in stop band, and cut off frequency at 4.7 kHz). Figure 94 shows the frequency response of the two-channel filter bank. Figure 95 reports the FRFs of the hybrid-method measured RIRs.

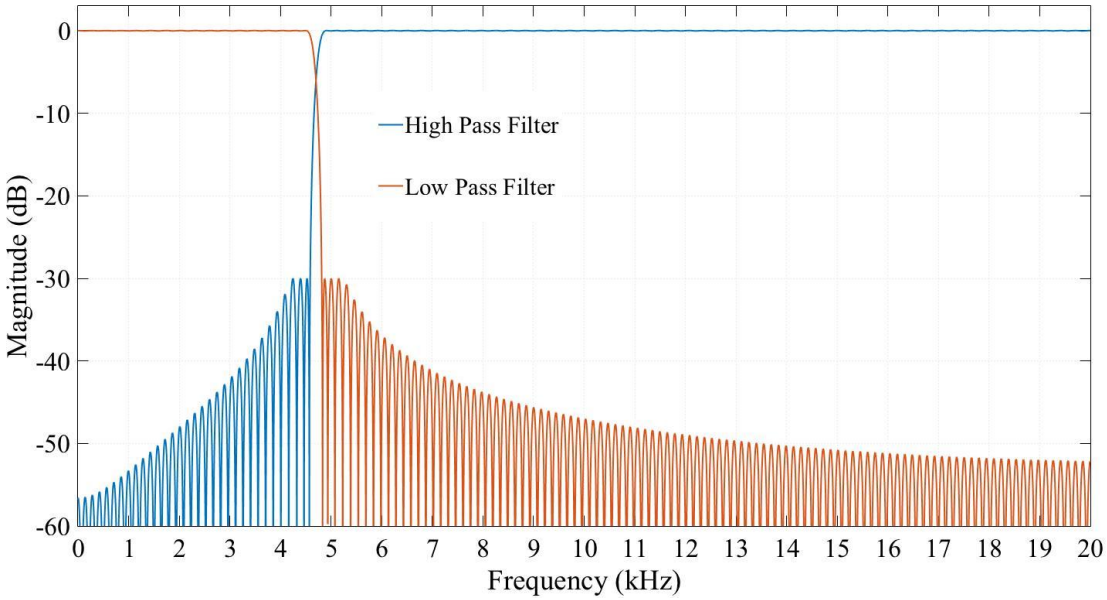


Figure 94 – The two-channel filter bank used to perform the time alignment of the RIRs

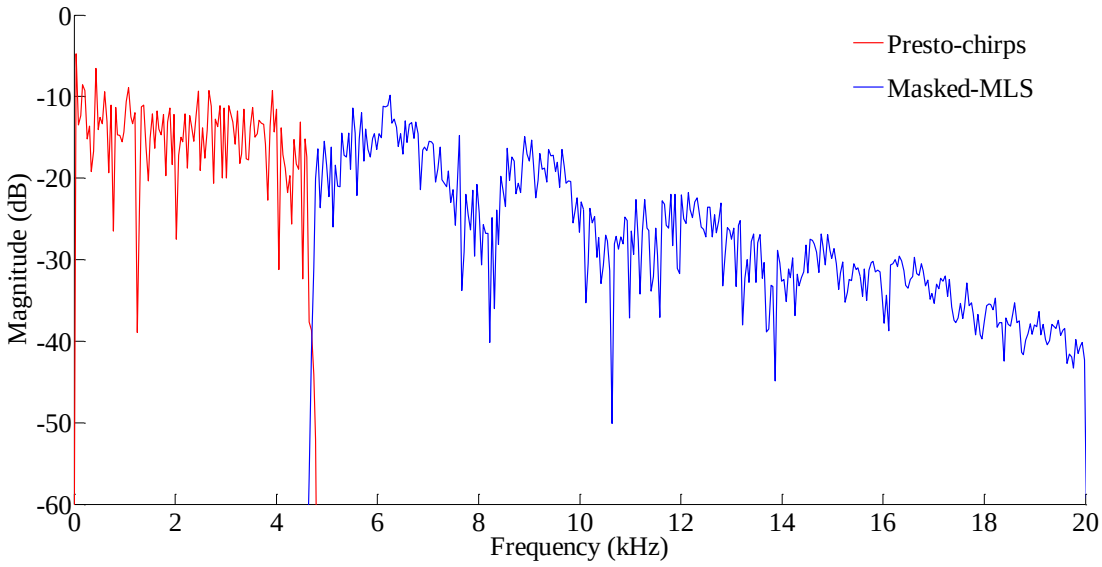


Figure 95 - FRFs of the band limited RIRs measured using the presto-chirps and masked-MLS, after the filtering process using the two-channel filter-bank.

Figure 96 shows the reference band limited RIRs. Those RIRs are compared to find time-amplitude relations between their waveforms. The positions (in sample) and the amplitudes of the higher peaks of the two waveforms are used to as references for the waveforms of the hybrid-measured RIRs. In practice, those latters, are time-shifted and energy normalises to let them match the reference RIRs.

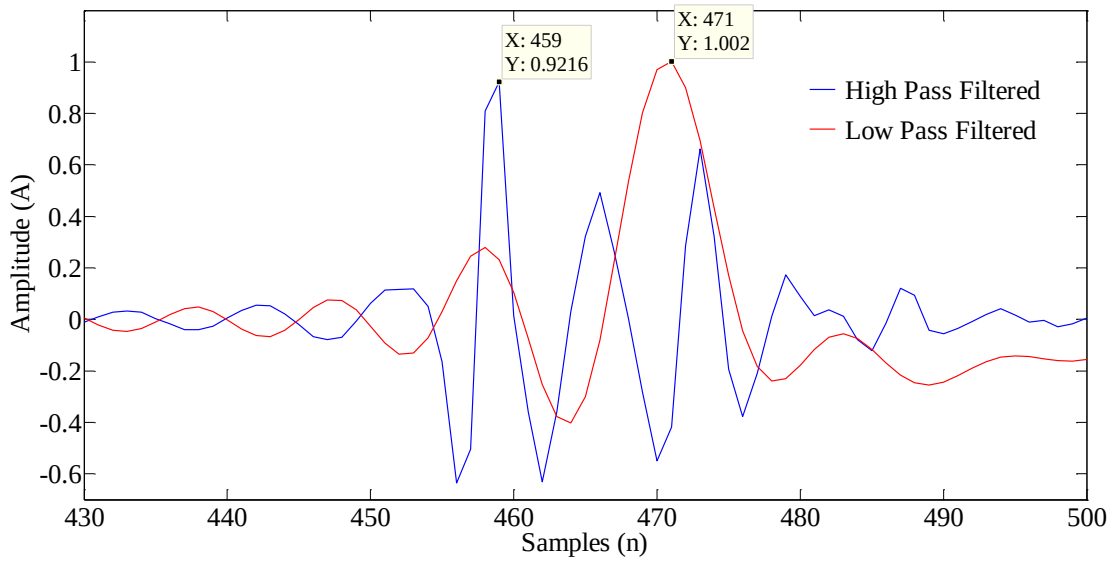


Figure 96 – Band limited RIRs obtained by filtering a reference broadband measurement through the two-channel filter bank.

7.3 Recombination Results

The hybrid-measured RIR is compared with two reference RIRs, to prove if the merging process is successful. Figure 97 reports the cross-correlation analysis of the RIR measured using the Hybrid method versus the RIRs measured using the MLS, and the ESS probes. The early part (up to 0.3 seconds) of the hybrid-measured RIR correlated quite perfectly with both the reference RIRs. This result can also be noted in the magnifications of the early peak of the RIRs shown in the yellow boxes within the figures. Differences between the RIRs appear minimal and are mainly relative to the magnitude of the peaks. Correlation decreases in the late part of the RIRs, due to the presence of noise. The reference ESS had lower SNR when compared with the reference MLS (which has been averaged several times), and this is the reason for the earlier degradation of correlation values, compared with the MLS case.

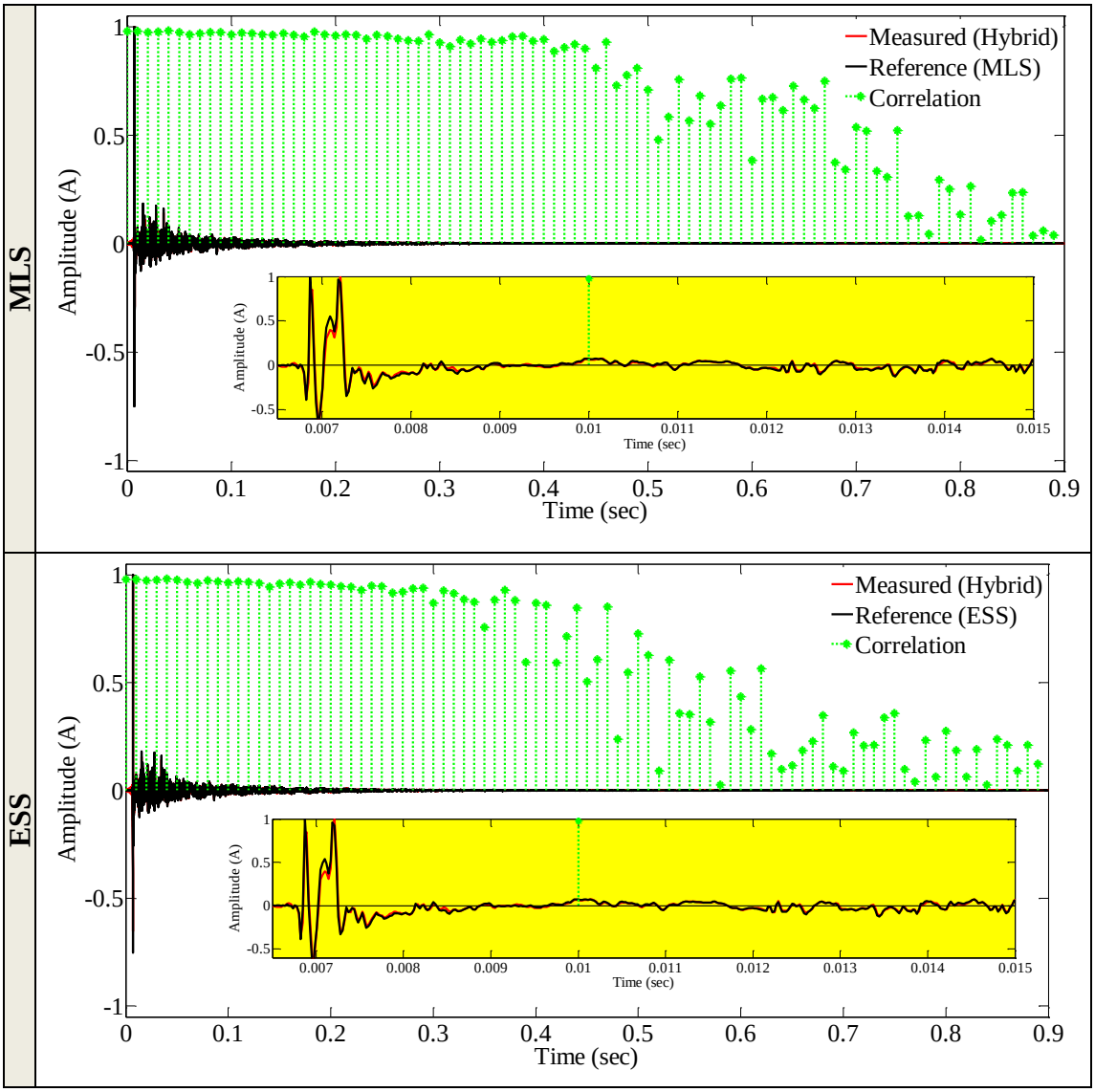


Figure 97 – Cross-correlation analyses of the RIR measured using the Hybrid method with the same RIR measured using an MLS probe (upper panel), and a ESS probe (bottom panel).

Figure 98 shows the coherence and SNR comparisons between the reference and measured RIRs. In the top panel, the comparison between the RIR measured using the Hybrid method versus the RIR measured using the MLS probe is reported. In the bottom panel, the Hybrid measurement is compared with the RIR measured using the ESS probe. These comparisons show that, on one hand there are not major anomalies in the separation point between the two methods (last vertical green line, which identifies the end frequency of the higher presto-chirps). On the other hand, they

highlight the presence of noise in the part of the spectra measured using the masked-MLS method, especially from frequencies above 8 kHz. In contrast, the part of the spectra measured by the presto-chirps method shows high coherence with the reference RIRs, denoting the reliability of the method.

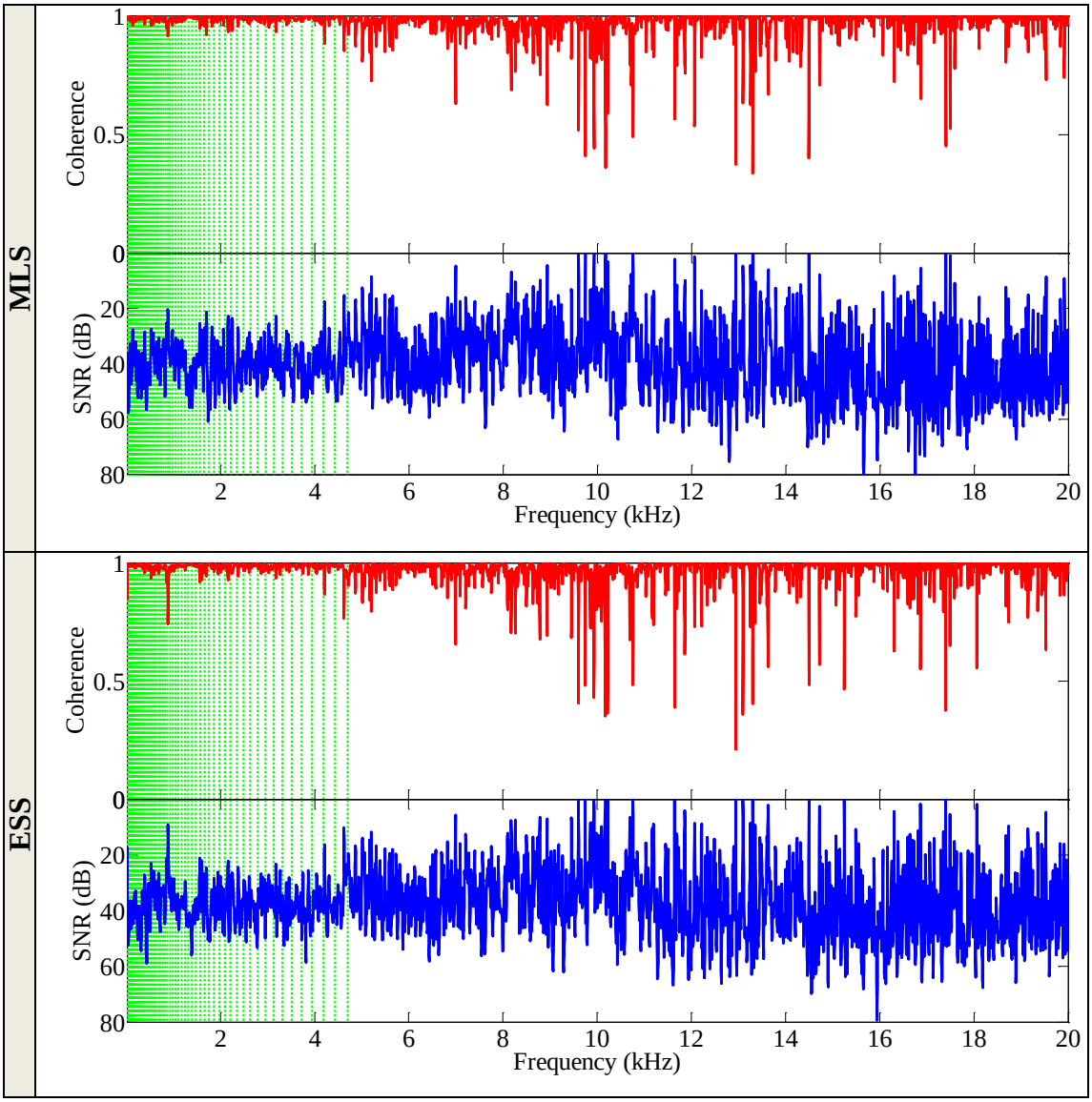
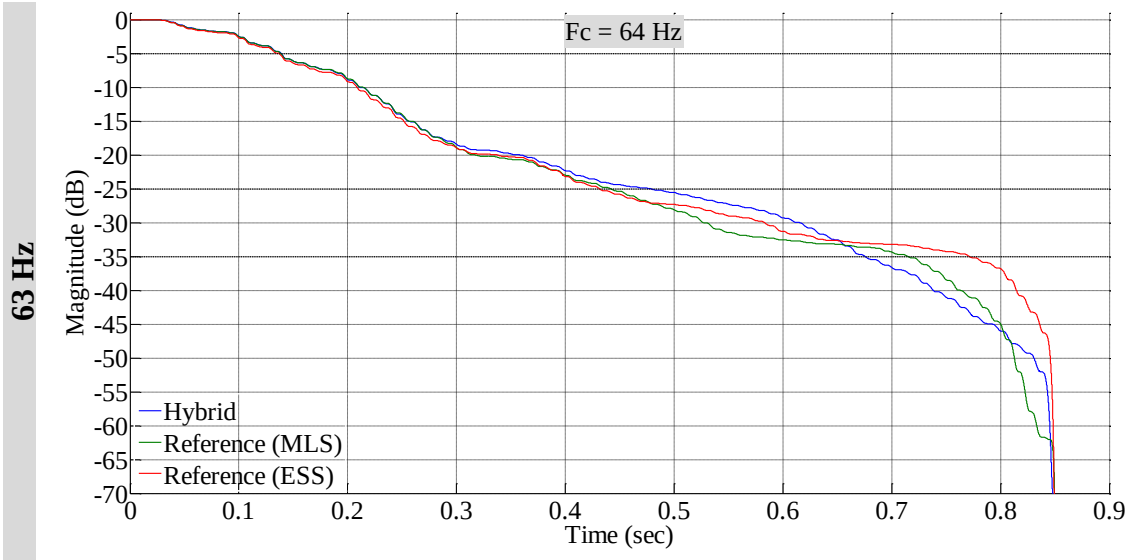


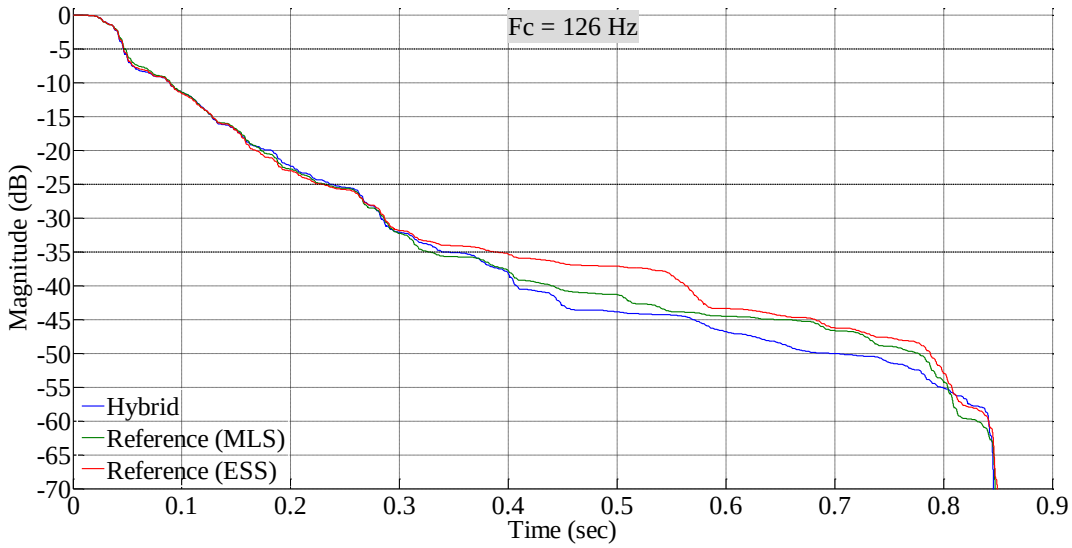
Figure 98 – Coherence and SNR analyses of the RIR measured using the Hybrid method with the same RIR measured using an MLS probe (upper panel), and a ESS probe (bottom panel).

Figure 99 reports the EDCs (in the octave bands 63 Hz – 16 kHz) calculated from RIRs measured using the Hybrid, the MLS and the ESS methods, as labelled. At the

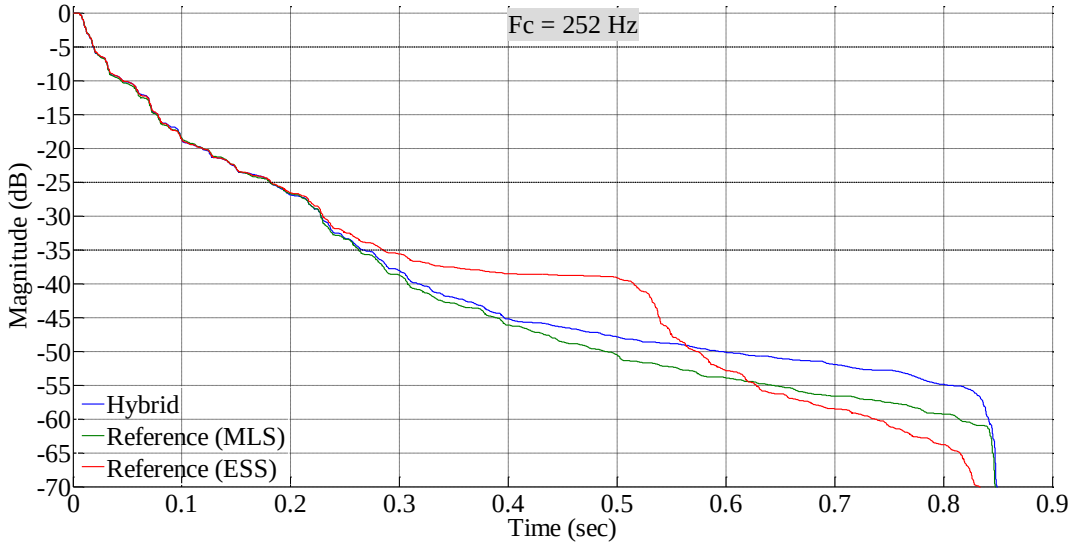
lower sub bands (63, 125 and 250 Hz) the EDCs have similar patterns up to about –35 dB. At the 250 Hz sub band the reference ESS suffered from some noise disturbance. At this same band, the EDCs of the Hybrid and MLS match closely, although the former has a slightly shorter noise-free decay range. For the 500, 1k and 2k Hz sub bands the Hybrid method, which is in the range of the presto-chirps method, equals both the reference EDCs reaches achieving even more than 50 dB of noise-free decay range. Starting from the 4 kHz sub band, which includes the transition point between the two methods, the EDCs related to the Hybrid method are progressively more affected by noise as the frequency increases. The EDC related to the ESS probe show a smaller SNR compared with the other EDCs. In the 16 kHz band, the EDC of the hybrid-measured RIR bends considerably from the reference EDC of the MLS. This denotes the poor SNR achieved by the masked-MLS method in that band, which besides the higher used level, still cannot achieve the 50 dB target.



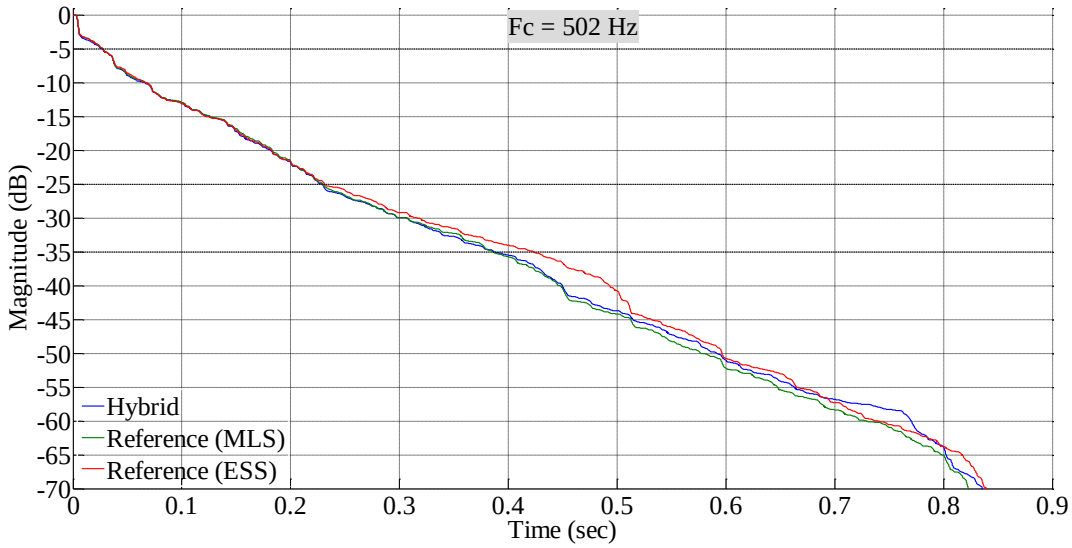
125 Hz



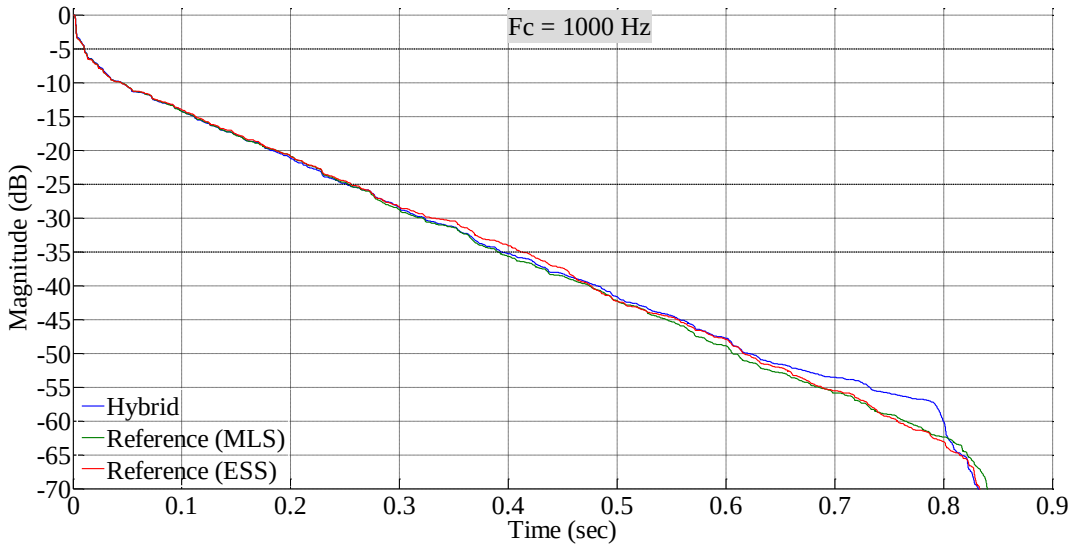
250 Hz



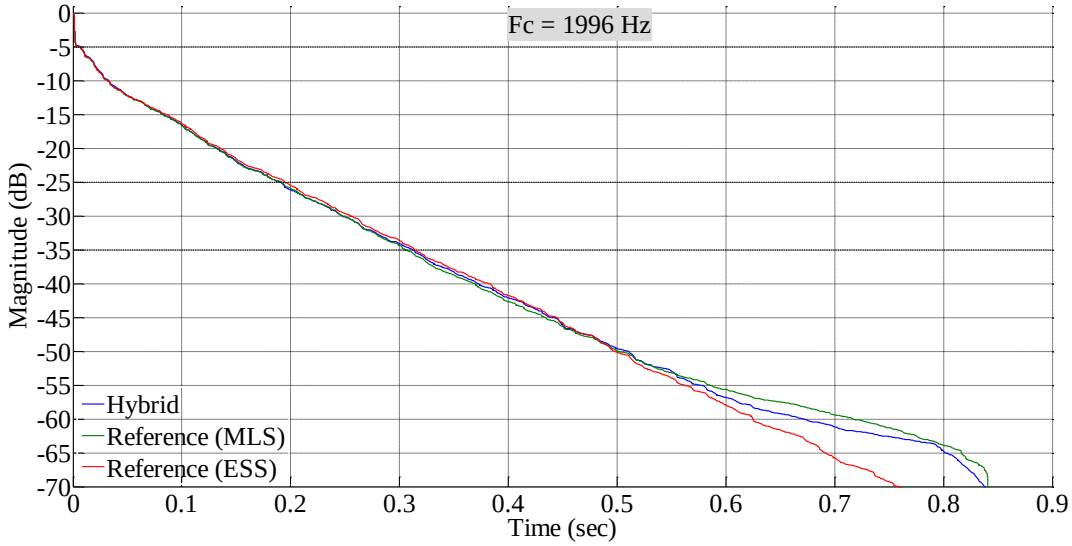
500 Hz



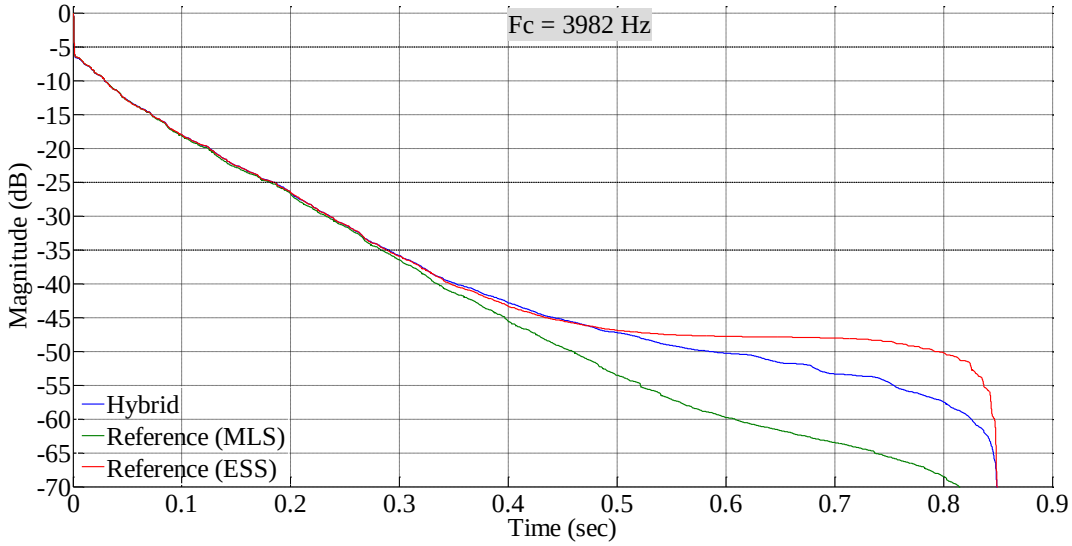
1 kHz



2 kHz



4 kHz*



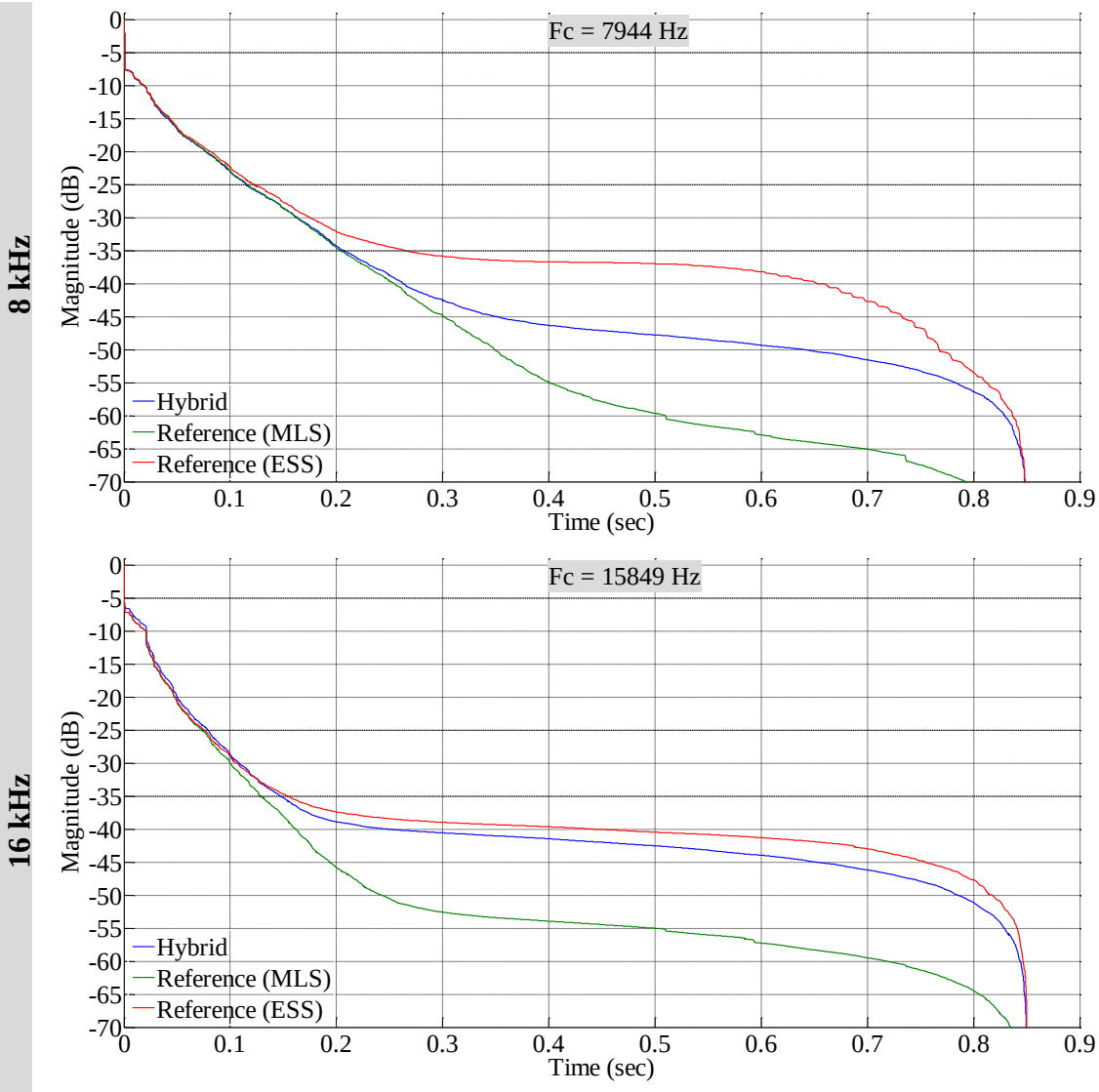


Figure 99 – EDCs calculated for the sub bands from 63 Hz to 16 kHz, of the RIR measured using the Hybrid method, an MLS and an ESS probe, as labelled.

Table 13 reports the acoustical parameters calculated from the reference MLS-measured RIR (labelled as "Ref."), and the difference within the parameters calculated from the Hybrid-measured RIR (labelled as "Diff."). The parameters are shown in the octave band from 63 to 16 kHz. The Hybrid method resolves to calculate acoustic parameters within difference limens with the reference values in all the sub bands, in the octave bands 63 and 125 Hz, for the RT30 parameter. At the 16 kHz sub band, the Hybrid method SNR (measured by the masked-MLS method) was too low

to lead to an accurate estimation of the acoustic parameters. The acoustical parameters have been estimated using the plug-in Aurora.

Parameters	**C80 [dB]** *(DL = 1 dB)*		**D50 [%]** *(DL = 5%)*		**EDT [sec]** *(DL = 0.024)*		**RT30 [sec]** *(DL = 0.024)*	
Bands [Hz]	**Ref.**	**diff.**	**Ref.**	**diff.**	**Ref.**	**diff.**	**Ref.**	**diff.**
63	5.52	0.15	58.76	0.94	0.82	0.00	0.81	0.15
125	12.00	0.03	87.08	1.30	0.34	0.01	0.59	0.02
250	16.00	0.66	92.60	1.23	0.28	0.02	0.54	0.01
500	11.78	0.16	87.39	0.31	0.46	0.02	0.77	0.01
1k	12.45	0.02	91.45	0.14	0.35	0.01	0.81	0.01
2k	15.66	0.04	94.96	0.06	0.29	0.01	0.64	0.00
4k	15.72	0.36	94.72	0.38	0.41	0.01	0.61	0.01
8k	21.61	0.38	98.31	0.61	0.26	0.02	0.44	0.01
16k	27.02	3.87	99.25	0.98	0.29	0.04	0.26	0.03

Table 13 – Reference acoustical parameters calculated from a RIR measurement using an MLS probe are reported in the columns labelled "Ref.". The columns labelled "Diff" reports the difference between the reference values and the values calculated from the RIR measured using the Hybrid method. The parameters are calculated using the plug-in Aurora.

Figure 100 reports a visual comparison of the (smoothed) FRFs of the aforementioned RIRs (1 Hybrid, 2 references). The Hybrid-measured FRF does not match the reference FRFs for the frequencies above 16 kHz, and this was due to the low SNR achieved by the masked-MLS method.

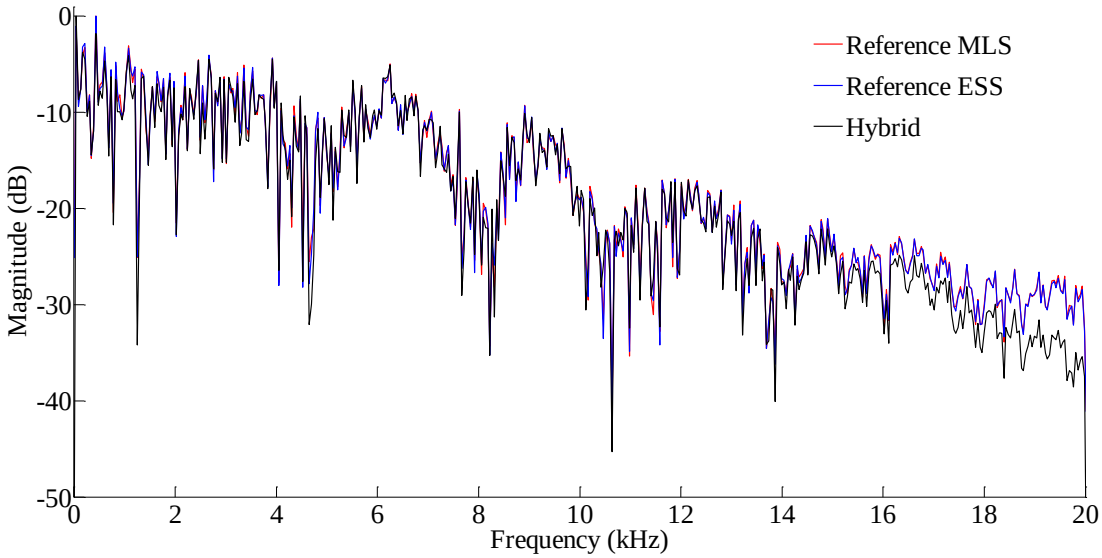


Figure 100 – Comparison of the FRFs calculated from the RIRs measured using the Hybrid method, an MLS and an ESS probes.

7.4 Discussions

A band limited RIR, measured using a presto-chirps probe of 60 dB L_{Aeq}, has been successfully re-aligned with another band limited RIR, measured using a masked-MLS of 50 dB L_{Aeq}. Such alignment has been manually performed by using a known broadband RIR as a reference.

The recombined RIR correlates well with the reference MLS-measured RIR, for the first 0.4 seconds. Noise becomes predominant afterwards, causing a progressive decrease of correlation. The presence of such high noise in the measured RIR is due to the low SNR achieved by the masked-MLS method at high frequencies. The use of a measurements level of 50 dB L_{Aeq} for the masked-MLS probe, which is higher than the levels used in chapter 6, was to experiment if whether, at least with such level, the masked-MLS would have worked. Indeed, the masked-MLS, gave accurate estimations of the tested acoustical parameters in the 4 kHz and 8 kHz octave bands,

which are of interest in room acoustics. The SNR was instead too low in the 16 kHz band to give accurate results. The presence of noise in that band has been pointed out comparing the FRFs of the hybrid-measure RIR with the reference RIRs. This limitation of the masked-MLS method at high frequency could be solved, as already discussed, by using a forward filter that will also account for the decrease of hearing sensitiveness at high frequency.

The masked-MLS should be redefined taking into account the found limitations. For example, it could be possible by using a forward filter with a roll-off lower than the 25 dB/octave used in this thesis, to perform accurate measurements even using a level of the probe minor than the 50 dB L_{Aeq}, used in this validation. By doing this, the MLS will probably be noticeable, because it will be above the masking threshold. However, it may results not obtrusive, and therefore usable for occupied measurements.

On the other hand, by using a presto-chirps stimulus of 60 dB L_{Aeq}, it has been proven that the method can still lead to accurate results. Using low-level probes is a straightforward way to reduce measurements annoyance. The presto-chirps method, supported by the use of low-level music-like stimuli, is reckoned a workable solution to address occupied measurements.

8 Conclusions and Further Work

To address the problem of occupied impulse response measurements a “Hybrid” method, which consists of two complementary measurement methods, namely the presto-chirps and the masked MLS. Each method supplies the impulse response of a part of the audio spectrum, which are subsequently merged together to give a broadband impulse response.

Specifically:

- The presto-chirps method is used to measure impulse responses in the frequency range from 27.5 Hz to 4.69 kHz. It utilises windowed semitone-width linear frequency varying chirps, i.e. the presto-chirps, to mimic musical tones. This method represents a paradigm of designing probe stimuli that meet both aesthetic and scientific requirements. It has been proven reliable in performing impulse response measurements. However, the used stimulus lacks the musicality that could make impulse response measurements unobtrusive.
- The masked-MLS method is used to measure impulse responses in the frequency range from 4.69 kHz upwards. It proposed to use a low-level MLS probe masked by a low-pass filtered music to perform non-intrusive measurements. This method has been proven unreliable in determining impulse responses, for the tested scenario and levels of the music and MLS. A redefinition of the method is suggested, which could increase the accuracy of measurements. However, the feasibility of inaudible measurements should be further assessed through psychoacoustic tests.

The presented hybrid method was primarily developed for large venues like auditoria, halls, theatres and cinemas, where an audience has a significant impact on their acoustics. However, it may also be used to analyse the acoustics of smaller enclosures like living rooms, cars, cabins, classrooms, and offices. Testing of public address systems is another potential application.

8.1 The Presto-Chirps Method

The presto-chirps method uses narrowband linear chirps as stimuli to measure room impulse responses. It represents a novel approach to in-situ measurements in that the presto-chirps are centred on chromatic scales and therefore they are effectively musical notes that can be used to compose musical stimuli. This approach supplies a less obtrusive alternative to the annoying sound of broadband chirps, while being able to perform accurate measurements of RIRs under occupied conditions. In fact, presto-chirps allows the use of any musical structure, e.g. scales and chords, if no frequency overlapping happens in a time shorter than the used analysis window. Frequency decreasing presto-chirps can also be used. This solution is deemed valuable in tackling the problem of stimuli obtrusiveness. However, truncating (time windowing) continuous chirp signals into many small segments is known to be a challenging task. Brick-wall type windows produce extensive frequency ringing in the spectra of the measured transfer functions, which, if not reduced, will bias the measurements and compromise the attainable SNR. One of the contributions of this research is to identify suitable windowing techniques to mitigate the fade in and fade out artefacts when segmenting chirps. This is successfully tackled by applying a Tukey window (tapered cosine) to the presto-chirps. An important feature of the Tukey window is

that its shape can be varied, from a rectangular to a cosine (Hanning), by simply changing a parameter of the window.

When viewing the presto-chirps as filters, applying a Hanning window to presto-chirps produces a filter with narrow pass bands but wider transition bands, while the opposite is obtained when a rectangular window is used. The wider the transition bands and the more the presto-chirps are frequency spaced to one another. This produces discontinuities in the excited spectrum, resulting in a comb filtering of measured RIRs. To avoid this, a variable frequency overlapping technique has been developed, which consists of first changing the presto-chirps start and stop frequencies by an amount related to the used Tukey window, and then windowing. Nonetheless, even though the overlapping technique effectively makes the presto-chirps frequency response continuous over the whole audio spectrum, it does not produce a flat response. Ripples are present at those frequency points that correspond to the start and stop frequencies of presto-chirps, which hinder accurate measurements of impulse responses. Nonetheless, it has been proven that using a Hanning-window on presto-chirps can leads to a bias in the frequency domain limited to ±1 dB (at the low end of the audio spectrum), which decrease to ±0.5 dB when presto-chirps with a length greater than one second are used. Moreover, the introduced error in the time domain results is sufficiently small enough to allow correct measurements of impulse responses.

A limitation on the use of presto-chirps is the risk that frequency-overlapping presto-chirps can produce “deconvolution residuals” along measured RIRs. However, this risk is limited to those presto-chirps belonging to a same block of deconvolved audio,

besides the fact that positions and shapes of residuals depend on the time-frequency characteristics of the overlapping presto-chirps, which may therefore not create residuals within the length of the RIRs.

Although it might be analytically difficult to predetermine if a particular stimulus will produce biased measurements, it suffices calculating the autocorrelation function of a stimuli to know where residuals would appear in the measured RIRs. Stimuli can then be changed accordingly. However, since deconvolution has to be performed in blocks of audio related to the length of the impulse response to be measured, specific stimuli can only be used to measure impulse response up to a determined length. This length is resolute by the distance between the time lag of peak of the ACF of a stimulus and the first appearing residual.

To validate the method, both computer simulations and in situ measurements have been carried out. For the in situ validations, a stimulus was created based on an actual song. With this stimulus and for the tested scenarios the presto-chirps method has been able to measure impulse responses within negligible errors when the presto chirps where Hanning-windowed. It has also been proven that the high SNR at low frequency bands is due to a large number of presto-chirps in a relatively small portion of the audio band. The possibility to arbitrarily define the amplitude, duration, and repetitions of each presto-chirp within stimuli is deemed useful to deal with noisy sub bands. Psychoacoustic tests are needed to determine the reduced annoyance of presto-chirps stimuli compared with broadband chirp probes. However, in the presented work this advantage has not been tested. The reason lies in the acknowledgement that more work should be done to create stimuli having an appreciable musicality.

8.2 The Masked-MLS Method

The masked-MLS method has been devised to be complementary to the presto-chirps method, to extend occupied acoustic measurements over the whole audio spectrum. The presented method relies on the masking effect produced by masker music to attempt at making a low-level MLS probe inaudible. The accuracy of MLS is strictly related to the SNR that can be achieved within the measurements, and even though it could theoretically be increased indefinitely through averaging, time variance is likely to limit such an increase. A high initial SNR allows to reduce the number of averages needed to bring the SNR up to sufficiently high values. For this reason, the masker music is low pass filtered (fc = 4.7 kHz), to maximise the SNR in the frequency region higher than that cut off point. This ultimately shortens the measurement times, hence limiting possible detrimental effects associated with time variance. On the other hand, to favourite the masking of the MLS probe, this latter is frequency-shaped according to the masking threshold preventively calculated from the masker music, using the psychoacoustic model implemented in the MPEG1-layer 1 codec.

To define the level of the MLS by which it would result inaudible, in conjunction with the psychoacoustic model, informal listening tests have been carried out. It has been found that (based on the participants to the tests) the "inaudibility" level is of 28 dB L_{Aeq}, and that the "unobtrusiveness" level is of 37 dB L_{Aeq}. This for the used MLS frequency shape and room. In situ measurements have been carried out using a purposely built measurements chain, which involved the use of a MLS measurement software and of a digital audio workstation. Results of the measurements have highlighted the unfeasibility of MLS measurements performed with the

aforementioned levels. Chiefly, the SNR resulted too low in the higher frequency band to give reliable estimations of several acoustic descriptors.

It has therefore been concluded that the method needs to be redesigned to account for several founds inaccuracies that tainted the presented method. Mainly:

- The shaping of the MLS should account for the frequency responses of the room and of the loudspeaker used in the validations, to match more precisely the estimated masking threshold.
- The MLS frequency shaping could include a level raise at the higher part of the audio spectrum following the masking thresholds, which has been wrongly excluded in the used probe.
- A different, and newer, psychoacoustic model might be used to achieve more accurate masking thresholds.
- Different and more efficient masker music could be elaborated.
- Formal psychoacoustic tests must be done to determine precisely, with statistic relevance, the average and the not-disturbing masking threshold levels of a reformulated MLS.

8.3 Discussions and Further Work

The work has addressed the research question(s) as outlined in Chapter 1 and pursued what was stated in the original proposal: "*The project will develop a new method that employs purposely synthesized musical tone like stimuli and music masked test signals to replace the noisy test signals to enable occupied measurements*". This represents the first attempt to use narrow band chirps centred on musical notes, and masked MLS probes, to determine room impulse responses in the associated frequency range to accuracy adequate for the measurements of common acoustic parameters.

The Presto-chirps method employs a set of 89 presto-chirps, where each provides the measure of a semitone-width band of frequency, to measure impulse responses in the range from 27.5 Hz to 4.69 kHz. Presto-chirps are Tukey windowed to avoid ripples in their frequency domain. Moreover, a frequency-overlapping technique is used to mitigate frequency discontinuities. The shape of a Tukey window can be varied from a rectangular window to a Hanning window, by changing parameter in the equation of the window. Five different configurations have been tested, which corresponds to the values, 0 (rectangular), 0.25, 0.5, 0.75, 1 (Hanning). Presto-chirps can be freely arranged to compose probe stimuli, and this flexibility has allowed using a music-sounding stimulus, which has been derived from an actual song, to perform in situ measurements. Nonetheless, further work has to be made to compose entertaining musical stimuli. However, such refinement was beyond the scope of this thesis and it is left, as well as the experimenting of different centre frequencies and synthesis techniques, as future work.

In addition to the theoretical and mathematical description of the associated algorithms used to obtain impulse responses, computer simulations and real room measurements have been carried out to evaluate and validate the methods. Results have been presented in this thesis. For the tested scenarios, it has shown that when presto-chirps were Hanning-windowed, it was possible to achieve accuracy comparable with standard broadband measurement techniques. It has also been found that a Hanning-windowed presto-chirp stimulus can achieve a high SNR ratio, and this could contribute, along with the use of refined musical stimuli, to ease occupied measurements.

An additional feature of presto-chirps is that they can be arranged to form random sequences, therefore they may be frequency pre-emphasised to achieve diverse measurement goals. For example, in the presence of a narrowband background noise, a number of presto-chirps around the frequency range of the noise can be used and results averaged to achieve a generally equal SNR across the whole spectrum. Presto-chirp stimuli are also good candidates for loudspeaker measurements, since similarly to chirp based measurements, they can separate the non-linear part from the linear part in an impulse response and therefore handle systems with non-linearity. Hence, a set of presto-chirps covering the whole audio spectrum may be used in an electroacoustics context to measure loudspeakers and microphones. This would avoid technical workers being exposed to boring testing signals (i.e. broadband sine sweeps) for prolonged periods over their working hours. In addition, presto-chirps (musical) stimuli can be arranged to be representative to what is played through these transducers in real uses. For example, further applications of presto-chirps may be for the testing the impulse responses of the ear canals, to help equalising and setting

hearing aids. This procedure is nowadays commonly carried out using low-level MLS probes.

The masked-MLS method developed in this thesis extends the frequency range to higher frequency sub-bands, from 4.69 kHz to 16 kHz. The methodis derived from the psychoacoustic model developed for the MPEG-layer 1 codec It calculates a “lower” masking threshold from a selected piece of low pass filtered music to frequency-shape the MLS probe accordingly. Therefore the added MLS will be presented to occupants as a non-audible or barely audible hissing noise. However, investigation results have shown that the use of an inaudibility level of such frequency shaped MLS signal results in too low a SNR to allow accurate measurements within reasonable measurement durations. Using higher measurement levels gives better SNRs, which will therefore allow accurate measurements to be performed. In such a case, the hissing noise is not completely masked. Nonetheless, using music as a distracter can still facilitate performing unobtrusive measurements.

The work so far has indicated a few pathways for further work. Presto-chirps have been shown effective and sufficiently accuracy to be deployed to measure impulse responses or transfer function for room acoustics and electroacoustics. How to make a sequence of these chirplets more musical and how to compose the sequence that best compensate for background noise are interesting further work. As another significant aspect of further work, more sophisticated psychoacoustic and suitable models are sought, with the hope that they can help improve the masking effects and make the masked MLS completely non-audible.

References

Literature

Alrutz, H., & Schroeder, M. (1983). A fast hadamard transform method for the evaluation of measurements using pseudorandom test signals. In *Proceedings of the 11th International Conference on Acoustics* (pp. 235–238). Paris.

Aoshima, N. (1981). Computer-generated pulse signal applied for sound measurement. *The Journal of the Acoustical Society of America*, *69*(5), 1484–1488.

Baron Rayleigh, J. W. S. (1877). *The theory of sound* (p. 433). London .: Macmillan and co.

Bendat, J., & Piersol, A. (2011). *Random data: analysis and measurement procedures* (4th ed., p. 640). Wiley.

Bendat, J. S., & Piersol, A. G. (1980). *Engineering applications of correlation and spectral analysis* (2nd ed., p. 315). New York, NY, USA: Wiley-Interscience.

Beranek, L. (2004). *Concert Halls and Opera Houses* (2nd ed., p. 664). Springer-Verlag New York Inc.

Borish, J., & Angell, J. (1983). An efficient algorithm for measuring the impulse response using pseudorandom noise. *Journal of the Audio Engineering Society*, *31*(7), 478–488.

Bork, I. (2000). *A comparison of room simulation software — the 2nd round robin on room acoustical computer software*. *Acta Acustica*, vol. 86, pp. 943–956.

Bosi, M., & Goldberg, R. (2003). *Introduction to digital audio coding and standards* (p. 434). Springer-Verlag New York Inc.

Bradley, J. S. (1996). Optimizing the decay range in room acoustics measurements using maximum-length-sequence techniques. *Journal Audio Engeneering Society*, *44*(4), 266 – 273.

Bradley, J. S. (2011). Review of objective room acoustics measures and future needs. *Applied Acoustics*, *72*, 713–720.

Campanini, S., & Farina, A. (2009). A new audicity feature: room objective acoustical parameters calculation module. In *Proceedings of the Linux Audio Conference*. Parma (Italy) 16-19 April.

Chu, W. T. (1990). Impulse-response and reverberation-decay measurements made by using a periodic pseudorandom sequence. *Applied Acoustics*, *29*, 193–205.

Cook, G., & Zaknich, A. (1996). Theory and implementation of extensions to time delay spectrometry. In *Preprints of the 6th Australian Regional Convention of the Audio Engineering Society*. Melbourne.

Couvreur, L., Ris, C., & Couvreur, C. (2001). Model-based blind estimation of reverberation time: application to robust asr in reverberant environments. *INTERSPEECH*, *6*(2), 2635–2638.

Cox, T., Li, F., & Darlington, P. (2001). Extracting room reverberation time from speech using artificial neural networks. *Journal of the Audio Engineering Society*, *49*(4), 219–230.

Cremer, L., Muller, H. A., & Schultz, T. J. (1982). *Principles and Applications of Room Acoustics Volume 1: Geometrical, Statistical and Psychological Room Acoustics*. Applied Science.

Davies, W. D. T. (1966). Generation and properties of maximum length sequences. *Control*, 302–304, 364–365, 431–433.

Dunn, C., & Hawksford, M. (1993). Distortion immunity of mls-derived impulse response measurements. *Journal of the Audio Engineering Society*, *41*(5), 314–335.

Dunn, C., & Rife, D. (1994). Comment on " distortion immunity of mls-derived impulse response measurements"*. *Journal of the Audio Egineering Society*, *42*, 490 – 497.

Eyring, C. F. (1930). Reverberation time in "dead" rooms. *Journal of the Audio Engineering Society*, (1), 217–226.

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Preprints of the 108th Audio Engineering Society Convention*. Paris.

Farina, A. (2007). Advancements in impulse response measurements by sine sweeps. In *Preprints of the 122nd Audio Engineering Society Convention*. Vienna.

Farina, A., & Righini, F. (1997). Software implementation of an mls analyzer with tools for convolution, auralization and inverse filtering. In *Preprints of the 103rd Audio Engineering Society Convention*. New York, NY, USA.

Farnetani, A., Prodi, N., & Pompoli, R. (2008). On the acoustics of ancient greek and roman theaters. *The Journal of the Acoustical Society of America*, *124*(3), 1557–1567.

Fastl, H. (2013). Basics and applications of psychoacoustics. In *Proceedings of Meetings on Acoustics* (Vol. 19). Montreal: Acoustical Society of America through the American Institute of Physics.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and Models* (3rd ed., p. 463). Springer-Verlag Berlin Heidelberg.

Fitzroy, D. (1959). Reverberation formula which seems to be more accurate with nonuniform distribution of absorption. *The Journal of the Acoustical Society of America*, (31), 893–897.

Fletcher, H., & Munson, W. (1933). Loudness, its definition, measurement and calculation*. *Bell System Technical Journal*, *12*(4), 377–430.

Gade, A.C., Bradley, J.S., and Siebein, G.W., (1993). Effects of measurement procedure and equipment on average room acoustic measurements. *JASA*, 93(4), p. 2265.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1-2), 103–138.

Gockel, H., Moore, B. C. J., & Patterson, R. D. (2002). Asymmetry of masking between complex tones and noise: the role of temporal structure and peripheral compression. *The Journal of the Acoustical Society of America*, *111*, 2759–2770.

Gockel, H., Moore, B. C. J., & Patterson, R. D. (2003). Asymmetry of masking between complex tones and noise: partial loudness. *The Journal of the Acoustical Society of America*, *114*(1), 349.

Greest, M. C., & Hawksford, M. O. (1995). Distortion analysis of nonlinear systems with memory using maximum-length sequences. *IEE Proceedings - Circuits, Devices and Systems*, *142*(5), 345.

Hak, C. C. J. M., Wenmaekers, R. H. C., & van Luxemburg, L. C. J. (2012). Measuring room impulse responses: impact of the decay range on derived room acoustic parameters. *Acta Acustica United with Acustica*, *98*(6), 907–915.

Hall, J. L. (1997). Asymmetry of masking revisited: generalization of masker and probe bandwidth. *The Journal of the Acoustical Society of America*, *101*(2).

Harris, F. (1976). Windows, harmonic analysis and the discrete fourier transform. San Diego, CA, USA: Naval Undersea Center.

Harris, F. (1978). On then use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, *66*(1), 51–83.

Hee, J. (1990) Impulse response measurement using mls,193 – 205.

Hellman, R. (1972). Asymmetry of masking between noise and tone. *Perception & Psychophysics*, *11*(3), 241–246.

Helmholtz, H. (2009). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (p. 609).

Herlufsen, H. (1984a). Dual channel fft analysis (part i). *Brüel & Kjær Technical Review*.

Herlufsen, H. (1984b). Dual channel fft analysis (part ii). *Brüel & Kjær Technical Review*.

Heyser, R. (1967). Acoustical measurements by time delay spectrometry. *Journal of the Audio Engineering Society*, *15*(4), 370–382.

Hidaka, T., Nishihara, N., & Beranek, L. L. (1988). Relation of acoustical parameters with and without audiences in concert halls and a simple method for simulating the occupied state. *The Journal of the Acoustical Society of America*, *109*(3), 1028–1042.

Kendrick, P., Cox, T. J., Francis, F., Zhang, Y., & Chambers, J. (2007). Blind estimation of clarity, centre time and deutlichkeit from speech and music signals. In *Proceeding of the19th International Congress on Acoustics*. Madrid.

Kendrick, P., Cox, T. J., Li, F. F., Zhang, Y., & Chambers, J. a. (2008). Monaural room acoustic parameters from music and speech. *The Journal of the Acoustical Society of America*, *124*(1), 278–87.

Kirkeby, O., Nelson, P. a., Hamada, H., & Orduna-Bustamante, F. (1998). Fast deconvolution of multichannel systems using regularization. *IEEE Transactions on Speech and Audio Processing*, *6*(2), 189–194.

Li, F. F. (2005). Estimation of intelligibility from received arbitrary speech signals with support vector machine. In *Proceeding of the International Conference on Machine Learning and Cybernetics* (Vol. 6, pp. 3755–3760). IEEE.

Li, F. F., & Cox, T. J. (2001). Extraction of speech transmission index from speech signals using artificial neural networks. In *Preprints of the 110th Audio Engineering Society Convention* (pp. 1–5). Amsterdam.

Li, F. F., & Cox, T. J. (2003). Speech transmission index from running speech: a neural network approach. *The Journal of the Acoustical Society of America*, *113*(4), 1999.

Li, F. F., & Cox, T. J. (2007). A neural network model for speech intelligibility quantification. *Applied Soft Computing*, *7*(1), 145–155.

Li, F. F., & Serafini, M. (2010). Room impulse response measurement in the presence of non-trivial noise. In *Proceeding of Internoise*. Lisbon.

Li, K. M., & Lam, P. M. (2005). Prediction of reverberation time and speech transmission index in long enclosures. *The Journal of the Acoustical Society of America*, *117*, 3716–3726.

Lyon, R. F., Katsiamis, A. G., & Drakakis, E. M. (2010). History and future of auditory filter models. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (pp. 3809–3812). IEEE.

Marie, P., & Jeong, C. (2012). Audience noise in concert halls during musical performances. *The Journal of the Acoustical Society of America*, *131*(4), 2753–2761.

Meyer, J. (1993), Vibrato sounds in large halls. *presented at the Stockholm Music Acoustics Conf*.

Millington, G. (1932). A modified formula for reverberation. *The Journal of the Acoustical Society of America*, (4), 69–82.

Müller, S., & Massarani, P. (2001). Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society*, *49*(6), 443–471.

Nielsen, J. (1996). Maximum-length sequence measurement of room impulse responses with high-level disturbances. In *Preprints of the 100th Audio Engineering Society Convention* (Vol. 4267, p. 25). Copenhagen.

Nielsen, J. (1997). Improvement of signal-to-noise ratio in long-term mls measurements with high-level nonstationary disturbances. *Journal of the Audio Engineering Society*, (3), 1063–1066.

Novák, A., Simon, L., Kadlec, F., & Lotton, P. (2010). Nonlinear system identification using exponential swept-sine signal. *IEEE Transactions on Instrumentation and Measurement*, *59*(8), 2220–2229.

Nuttall, A. (1981). Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *29*(1), 84–91.

Oppenheim, A. V., & Schafer, R. W. (2009). *Discrete-Time Signal Processing (3rd Ed.)* (p. 1120). Prentice-Hall, Inc.

Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE*, *88*(4), 451–515.

Pan, D. (1995). A tutorial on mpeg/audio compression. *IEEE Multimedia*, *2*(2), 60–74.

Paulo, J. P., Martins, C. R., & Bento Coelho, J. L. (2009). A hybrid mls technique for room impulse response estimation. *Applied Acoustics*, *70*(4), 556–562.

Pick, H. L. (1989). Inhibiting the lombard effect. *The Journal of the Acoustical Society of America*, *85*(2), 894.

Plomp, R., & Bouman, M. A. (1959). Relation between hearing threshold and duration for tone pulses. *The Journal of the Acoustical Society of America*, *31*(6), 749–758.

Poletti, M. (1988). Linearly swept frequency measurements, time-delay spectrometry, and the wigner distribution. *Journal of the Audio Engineering Society*, *36*(6), 457–468.

Ratnam, R., Jones, D. L., & O'Brien, W. D. (2004). Fast algorithms for blind estimation of reverberation time. *IEEE Signal Processing Letters*, *11*(6), 537–540.

Ratnam, R., Jones, D. L., Wheeler, B. C., O'Brien, W. D., Lansing, C. R., & Feng, A. S. (2003). Blind estimation of reverberation time. *The Journal of the Acoustical Society of America*, *114*(5), 2877.

Ream, N. (1970). Nonlinear identification using inverse-repeatm sequences. *Proceedings of the Institution of Electrical Engineers*, *117*(1), 213–218.

Reichardt, W., Abdel, O. A., and Schmidt W., (1974). Defintion und Me_grundlage eines objektiven Ma_es zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikdarbietung. *Acustica,* vol. 32, pp. 126-137.

Rife, D., & Vanderkooy, J. (1989). Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society*, *37*(6), 419–444.

Sabine, W. (1922). Collected papers on acoustics. Harvard University Press.

Satoh, F., Mori, J., Nishii, T., & Tachibana, H. (2013). Impulse response measurement in public space using musical signal including swept-sine signals. In *Proceedings of Meetings on Acoustics* (Vol. 19). Montreal.

Satoh, F., Nagayama, M., & Tachibana, H. (2002). Influence of time-variance in auditorium on impulse response measurement. In *Proceeding of the Forum Acusticum*. Sevilla.

Schoukens, J. (1988). Survey of excitation signals for fft based signal analyzers. *IEEE Transactions on Instrumentation and Measurement*, *37*(3), 342–352.

Schoukens, J., Pintelon, R. M., Rolain, Y. J., & Member, S. (2000). Broadband versus stepped sine frf measurements, *49*(2), 275–278.

Schroeder, M. (1979). Integrated-impulse method measuring sound decay without using impulses. *The Journal of the Acoustical Society of America*, *66*(Ml), 497–500.

Schroeder, M. R. (1965). New method of measuring reverberation time. *The Journal of the Acoustical Society of America*, *37*(3), 409–412.

Schroeder, M. R., & Hill, M. (1984). Progress in architectural acoustics and artificial reverberation: concert hall acoustics and number theory. *Journal Audio Engeneering Society*, *32*(4), 194–203.

Shailer, M. J. & Moore, B. C. J., (1983). Gap Detection as a Function of Frequency, Bandwidth, and Level. *J. Acoust. Soc. Am.*, 74(2).

Serafini, M., & Li, F. F. (2011a). Impulse response measurement with chirp-lets and masked noise stimuli. In *Proceedings of the Institute of Acoustics* (Vol. 33). Dublin.

Serafini, M., & Li, F. F. (2011b). Synthesised musical stimuli to enable occupied room impulse response measurement. In *Proceedings of the Institute of Acoustics* (Vol. 33). Brigthon.

Serafini, M., & Li, F. F. (2011c). Feasibility of masked pseudo random sequences for occupied measurement in auditoria. In *Proceeding of the Forum Acusticum*. Aalborg.

Serafini, M., Li, F. F., & Cox, T. J. (2010). Occupied measurement of room impulse response - feasibility and limitations. In *Proceedings of Internoise*. Lisbon.

Serafini, M., (2015). Presto-chirps: using flexible discrete chirp-lets based stimuli to measure acoustic impulse responses. In *Proceedings of ICSV22*, Florence.

Seraphim, (1958). Untersuchungen über die unterschiedsschwelle exponentiellen abklingen raushbandimpulsen, *Acustica* (8) 280-284.

Sette, W. (1933). A new reverberation time formula. *The Journal of the Acoustical Society of America*, (4), 193–210.

Smith III, J. O., & Abel, J. S. (1999). Bark and erb bilinear transforms. In *IEEE Transactions on Speech and Audio Processing*.

Stan, G., Embrechts, J., & Archambeau, D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, *50*(4), 249–262.

Suzuki, Y. (1995). An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *The Journal of the Acoustical Society of America*, *97*(2), 1119.

Svensson, P., & Nielsen, J. (1999). Errors in mls measurements caused by time variance in acoustic systems. *Journal of the Audio Engineering Society*, *47*(11), 907–927.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, *1*, 155–182.

Thrane, N. (1979). Discrete fourier transform and fft analyzers. *Bruel & Kjaer Techical Review*.

Tsolias, A. & Davies, W. J., (2014). Difference limen for reverberation time in auditoria. In *Proceeding of the Forum Acusticum.* Krakow

Vanderkooy, J. (1986). Another approach to time-delay spectrometry. *Journal of the Audio Engineering Society*, *34*(7/8), 523–538.

Vanderkooy, J. (1994). Aspects of mls measuring systems. *Journal of the Audio Engineering Society*, *42*(4), 219 – 231.

Vér, I., & Beranek, L. (2006). *Noise and vibration control engineering: principles and applications* (2nd ed., p. 976). John Wiley & Sons.

Vesa, S., & Härmä, A. (2005). Automatic estimation of reverberation time from binaural signals. In *Proceedings of the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005)* (Vol. 3, pp. 281–284). Philadelfia.

Vetter, K., & Di Rosario, S. (2011). Expochirptoolbox: a pure data implementation of ess impulse response measurement. In *Proceedings of the Institute of Acoustics*. Brigthon.

Vorlander, M., & Kob, M. (1997). Practical aspects of mls measurements in building acoustics. *Applied Acoustics*, *52*(3/4), 239–258.

Zhang, Y., Chambers, J., & Kendrick, P. (2006). Blind estimation of reverberation time in occupied rooms. In *EUSIPCO - European Processing Conference*. Florence, Italy.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, *33*(2), 248.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America*, *68*(5), 1523–1525.

Standards and Guidance

BS EN ISO 3382-1:2009 - Acoustics — Measurement of room acoustic parameters — Part 1: Performance spaces.

BS EN ISO 3382-2:2008 - Acoustics — Measurement of room acoustic parameters — Part 2: Reverberation time in ordinary rooms.

BS ISO 226:2003 - Acoustics — Normal equal-loudness-level contours.

BS EN ISO 18233:2006 - Acoustics — Application of new measurement methods in building and room acoustics.

BS 8233:2014 - Guidance on sound insulation and noise reduction for buildings

BS EN 61260:1996 - Electroacoustics — Octave-band and fractional-octave-band filters.

ISO/IEC 11172-3:1993 - Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Audio.

Appendix

A. Discussion on the Used Musical Stimulus

The composition that has been used to perform the in situ measurements, as already introduced in Section 3.6.3, has been rearranged from an actual song. Thus, "via del Corso" of Tero-Pekka Henell. Figure 101 shows the "re-arranged" score, whilst Table 14 shows the values extracted from the first 19 notes of the MIDI file associated with the song, which are used in Eq. (32) to construct the presto-chirps. Specifically:

- The field "pitch" relates with the variable *i*
- The field "amplitude" relates with the variable A (scaled in the range to 1 – 0, by dividing it for 127)
- The fields "start" and "end" related with the parameter N (= ($End\ Time - Start\ Time \cdot fs$)

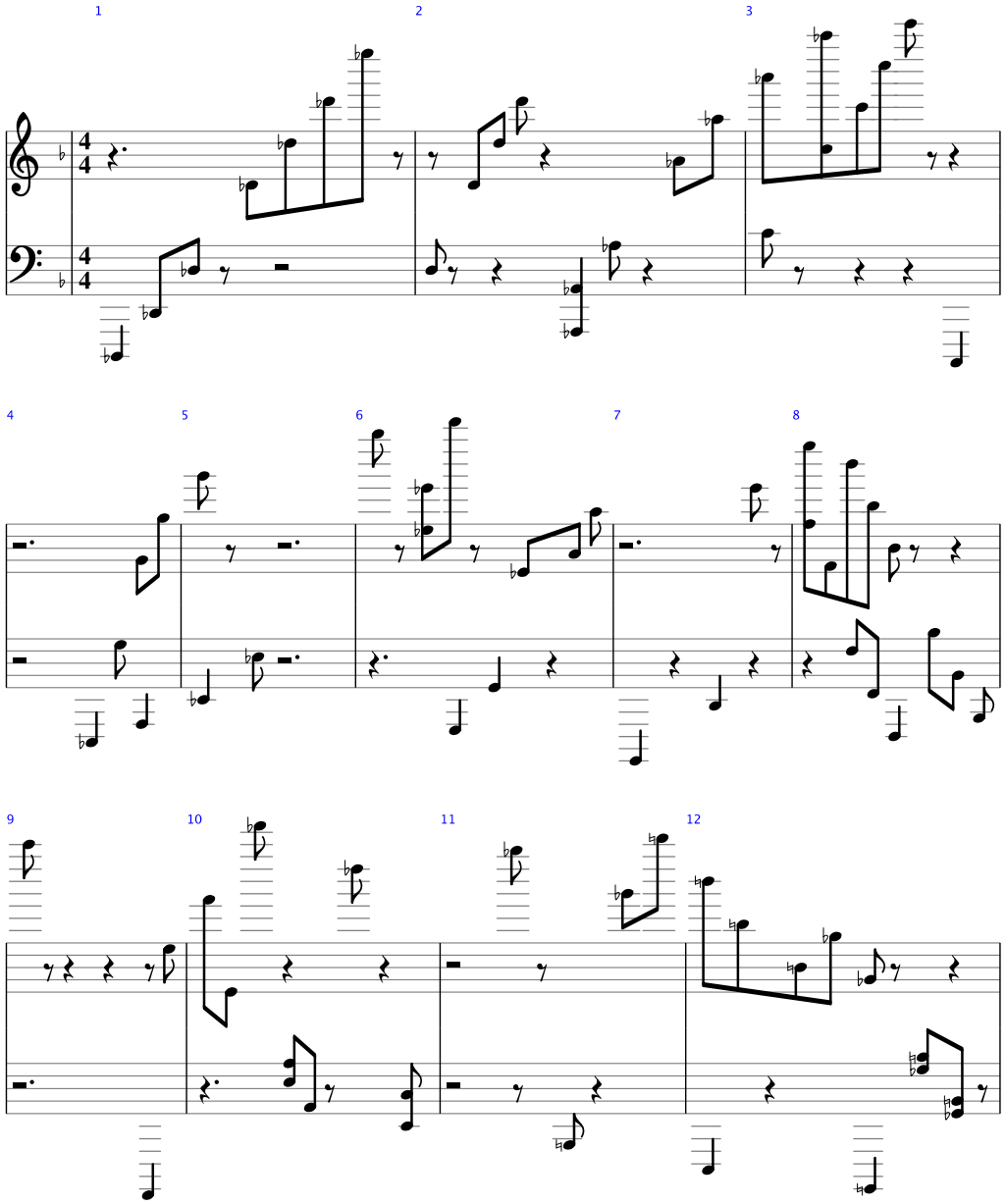


Figure 101 – Score of the stimulus used for the measurements, adapted from the song "Via del Corso" by Tero-Pekka Henell.

Pitch	Amplitude	Start Time (sec)	End Time (sec)
25	100	0	1.1
37	80	0.2	0.75
49	80	0.4	0.75
61	89	0.55	0.75
73	75	0.75	0.95
85	80	1	1.2
99	84	1.2	1.4
62	90	1.7	2
56	86	2	2.25
74	78	2	2.45

44	90	2.25	2.5
32	88	2.25	3
86	72	2.25	2.5
68	78	2.45	2.5
80	77	2.8	3
92	80	3	3.35
104	100	3	3.2
24	105	3	4.1
60	78	3	3.5

Table 14 – Table of parameters used to construct presto-chirps, extract from the MIDI file of the song used for the in situ measurements.

To check for the presence of notes that could lead to the generation of residuals, it is convenient to use a "piano-roll" to visualize the song (several music notation software have this visualisation mode). This representation is useful as it allows to see both the time distance (in seconds), and the frequency distance (in semitones) between notes. In fact, as mentioned, residuals are caused by the frequency overlapping of neighbouring presto-chirps simultaneously present in a same chunk of deconvolved audio. The length of this chunk corresponds to the maximum analysable length of a RIR. Figure 102 shows the piano-roll of the arranged song. The grey rectangles indicate when the keys of the keyboard shown in the bottom are played. The playing time is reported in the vertical bar on the left, in microseconds. The three notes circled with dashed lines indicate the "risky" situations. This is because notes are spaced by only one semitone, and played within about a second to one another. Practically, this time distance between semitones defines the limit of usability of this specific stimulus, and consequently the analysis window.

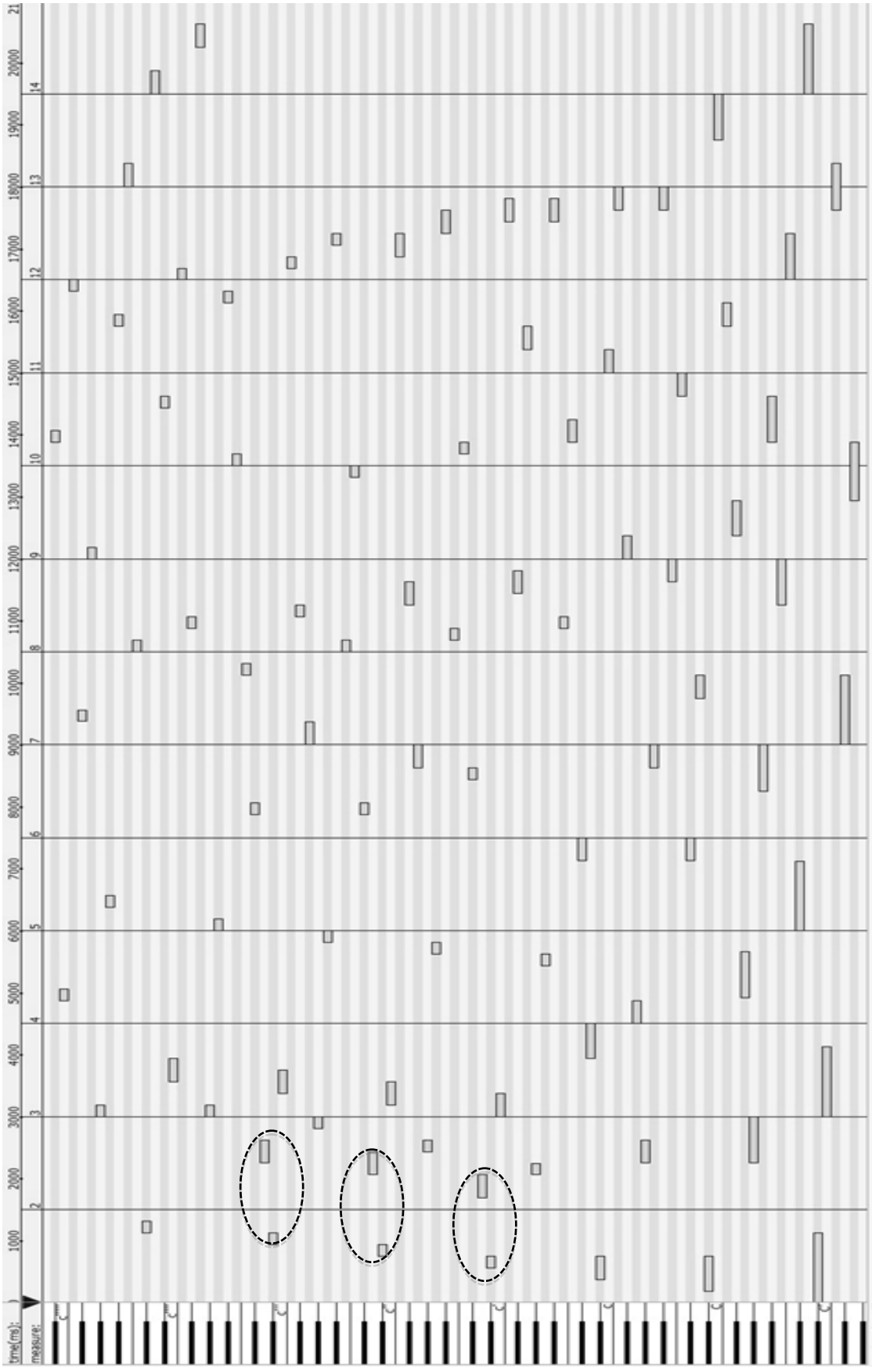


Figure 102 - Piano-roll representation of the adapted song "via del Corso".

As shown and discussed in Section 4.2.2.1, to calculate the EDCs (and then the acoustic parameters), longer analysis windows could be used to analyse specific sub bands. This is if residuals are not present in the filtered bands.

To shed light on this issue, a software simulation is used to determine the maximum allowed length of an RIR that can be measured with a stimulus (as suggested in Section 3.6.1). In this simulation, Hanning-windowed presto-chirps are used to play the aforementioned stimulus. For such cases, using an analysis window of 1.2 seconds results in the appearance of a residual in the late tail of the RIR. Figure 103 shows the comparison between a reference RIR and the one measured using the presto-chirp stimulus. The residual is also shown magnified in the yellow box within the figure.

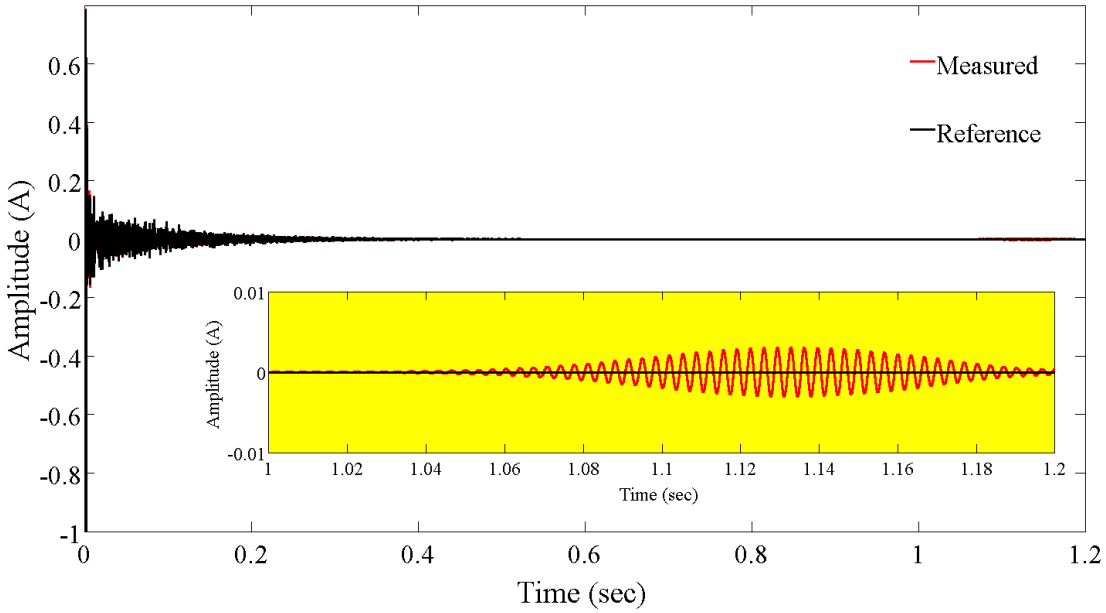


Figure 103 – Amplitude difference between the measured and reference impulse responses. The residual present in the late part of the RIR is shown magnified in the yellow box.

Figure 104 shows the magnitude differences between the reference RIR and the one measured using the presto-chirps method. It reveals the presence of an artefact at about 300 Hz, which is caused by the rightmost notes "circled" in the piano-roll. The residual is frequency limited because it is formed by the convolution of band limited chirps with a band limited "inverse" chirp. This justifies that band filtering the measured RIR is a workable solution to allow longer analysis windows to be used in those sub bands outside the frequency range of residuals. Note that acknowledging the frequencies of the residuals helps discovering the presto-chirps that have caused it. For example, in the case shown there were the presto-chirps with midi numbers 61 and 62 (notes D and D sharp – frequency 293.7 and 311.1 Hz).

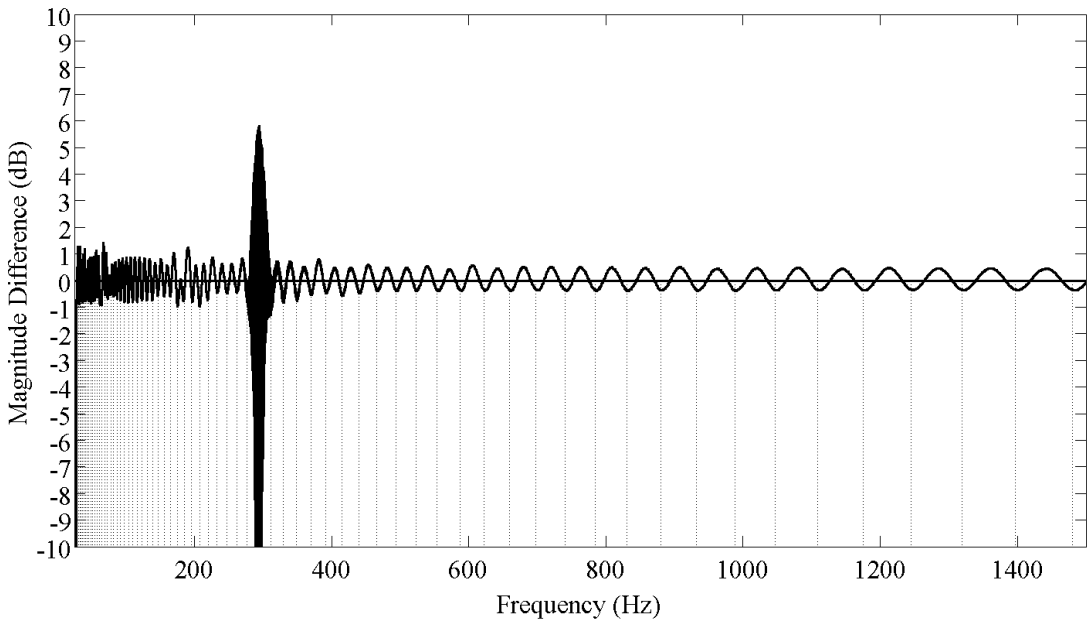


Figure 104 – Magnitude difference between the measured and reference frequency responses. The presence, and frequency content of the residual, is clearly visible.

For the sake of completeness, Figure 105 shows the magnitude differences between the reference and measured RIRs when the latter is analysed up to 2 seconds. The residuals correspond to the three circled notes shown in the piano roll. Figure 106

reports the comparison of the RIRs for such a case. The residuals are clearly visible along the tail of the RIR, magnified in the yellow box.

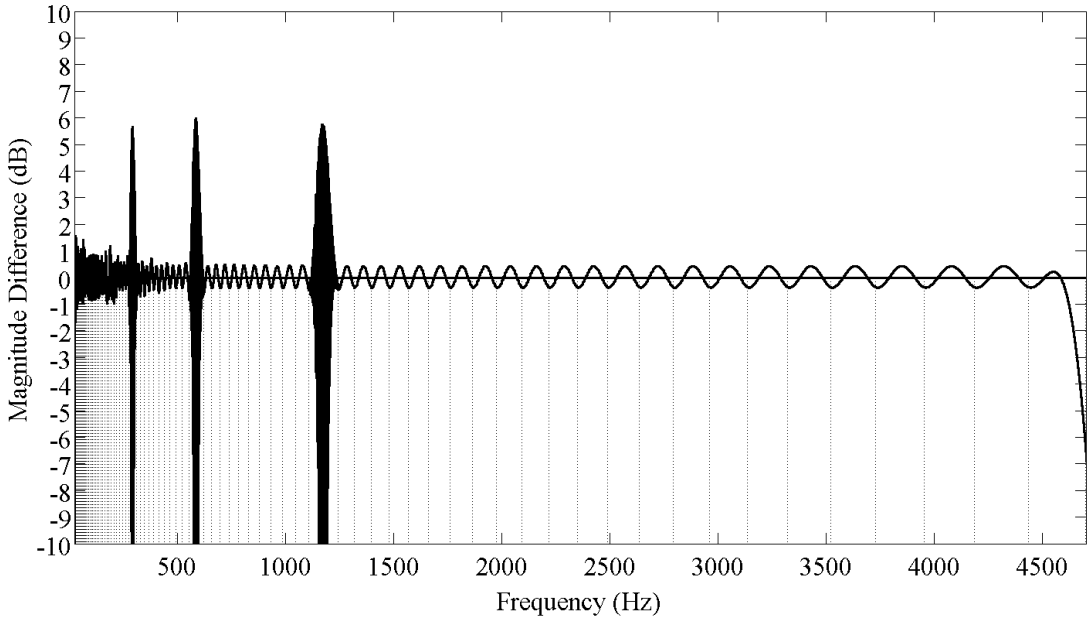


Figure 105 - Residuals in FRF of a RIR measured using a two second long AW.

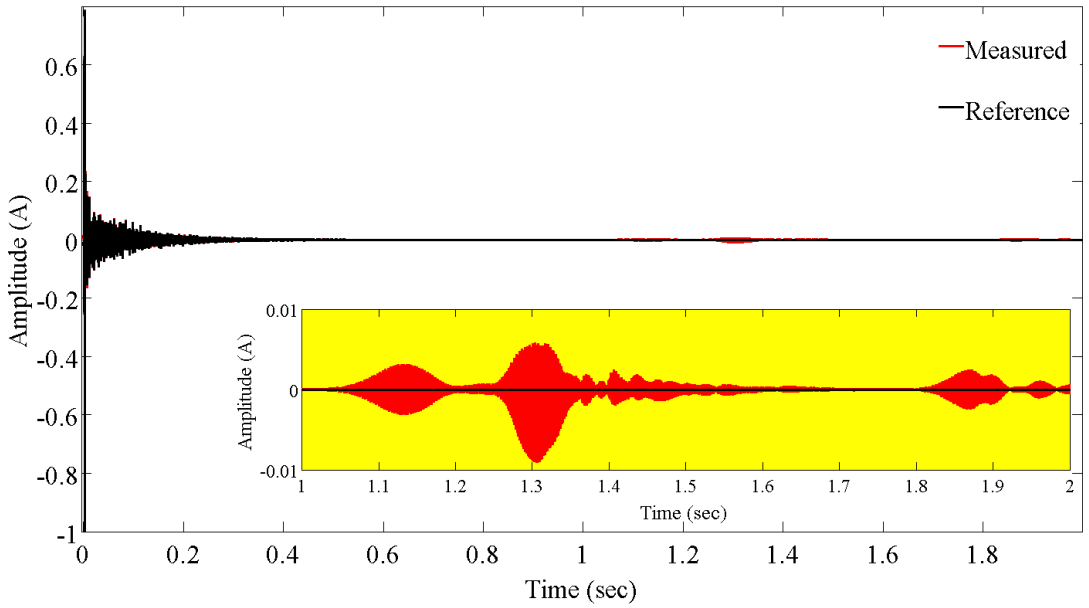


Figure 106 – Residuals in the tail of a RIR due to the use of a two seconds long analysis window.

B. Further Uses of Presto-chirps

Descending Presto-chirps

The LSS and ESS methods can only use chirps whose instantaneous frequency increases with time (ascending chirps). This is because distortions have a higher frequency than the source signal, and deconvolving chirps whose instantaneous frequency decreases with time (descending chirps), will repack the distortions in the causal part of the IR. On the contrary, since presto-chirps are band limited, distortions are filtered out during the deconvolution process, even if descending presto-chirps are used. Descending presto-chirps might contribute to the musicality of stimuli.

Multiband Analysis

Another effect that can be exploited using presto-chirps is the possibility of contemporary measuring several bands. This can reduce measurement times, while feeding more energy in the system compared with the use of broadband chirps. This can be conceptually equal to the difference between performing IR measurements using a (stepped) tone or a multitone signal. The latter can reduce measurements time and allows the analysis of IMD. The crest factor of a multitone signal can be made equal to that of a single tone by using in phase tones. Moreover, multiband analyses might be exploited to excite loudspeakers or other electronic devices with a signal that more closely mimics real musical sounds (or noise), compared with broadband chirps or single tones.

Multiband signals are also exploited in the composition of presto-chirps stimuli to form (presto-chirps) chords. Note that such chords might be subjectively felt slightly "out of tune".

The deconvolution of a multiband signal is carried out by using as many inverse filters as those that form the signal. However, when dealing with multi band signals a minimum "frequency" distance between presto-chirps must be preserved to avoid the formation of residuals. Note that a multiband stimulus has been used in Section 4.2.2 in the simulated measurement of the RIR of a church, which however was structured to avoid the formation of residuals (up to 2 seconds).

In the following is instead presented, for the sake of completeness, the measurement of an RIR by a multiband signal that generates residuals. This to show and discuss the limitations of a multiband analysis.

Figure 107 shows the stimulus used in this example. It is composed by two trains of one-second-long presto-chirps (both ascending and descending), followed by two seconds of silence. The presto-chirps in each train are separated by only a musical tone. Table 15 shows the first six notes of the MIDI file of the stimulus. Frequency is presented logarithmically. Such frequency separation is common in musical chords, although the same chords can be composed using different combinations, using more "spaced" notes. For example, some notes of the chords could be increased, or decreased, by an octave.

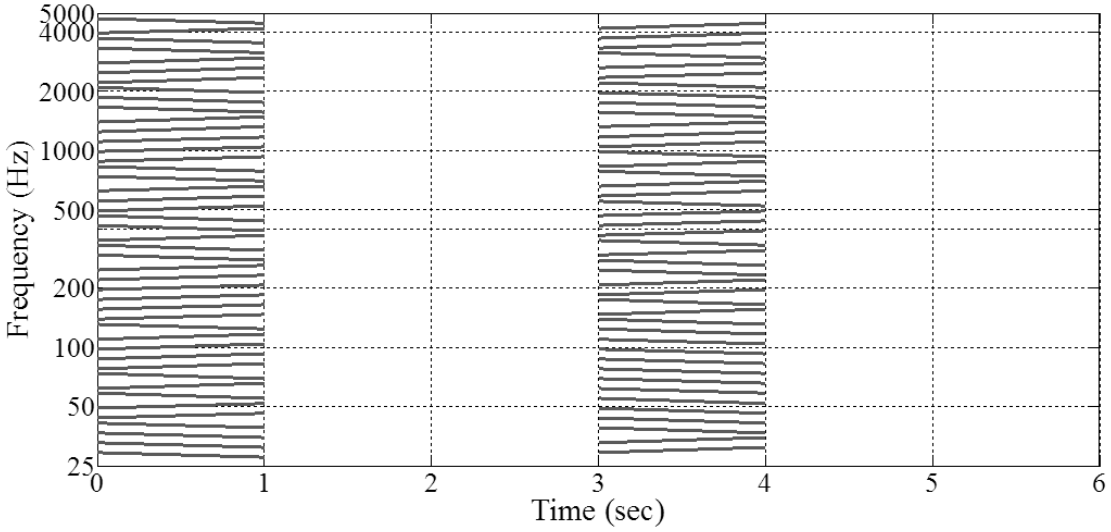


Figure 107 – Stimulus formed by 1 second long presto-chirps spaced by two seconds of silence. The presto-chirps in each group are spaced by two semitones.

Pitch	Amplitude	Start Time (sec)	End Time (sec)
21	127	0	1
22	127	3	4
23	127	0	1
24	127	3	4
25	127	0	1
26	127	3	4

Table 15 - First six notes of the MIDI file of the stimulus shown in Table 15.

Figure 108 shows the amplitude differences between the reference RIR and the RIR measured using the aforementioned stimulus for Hanning-windowed presto-chirps. As it can be seen, residuals are formed at around 0.8 seconds. This result highlights on the one hand the effect of overlapped frequencies and the production of residuals in the RIR (discussed in Section 3.3). On the other hand, it highlights that the usable frequency distance between presto-chirps in multiband stimuli depends on the length of the RIR that is measured. In fact, since the position of the residuals depends on the frequency difference of the instantaneous frequencies of the overlapped presto-chirps, thus on their duration and bandwidth, they could be made to appear far enough from the main peak of the IR, which could be time gated without affecting the results. In

general, shorter presto chirps produce residuals closer to the main peak, and vice versa. This time-frequency dependency between presto-chirps and the formation of residuals could be exploited to measure short IR (like the ones of loudspeakers, for example). Consequently, even frequency adjacent presto-chirps could be used providing they are sufficiently long enough to "push" the residuals far from the main peak.

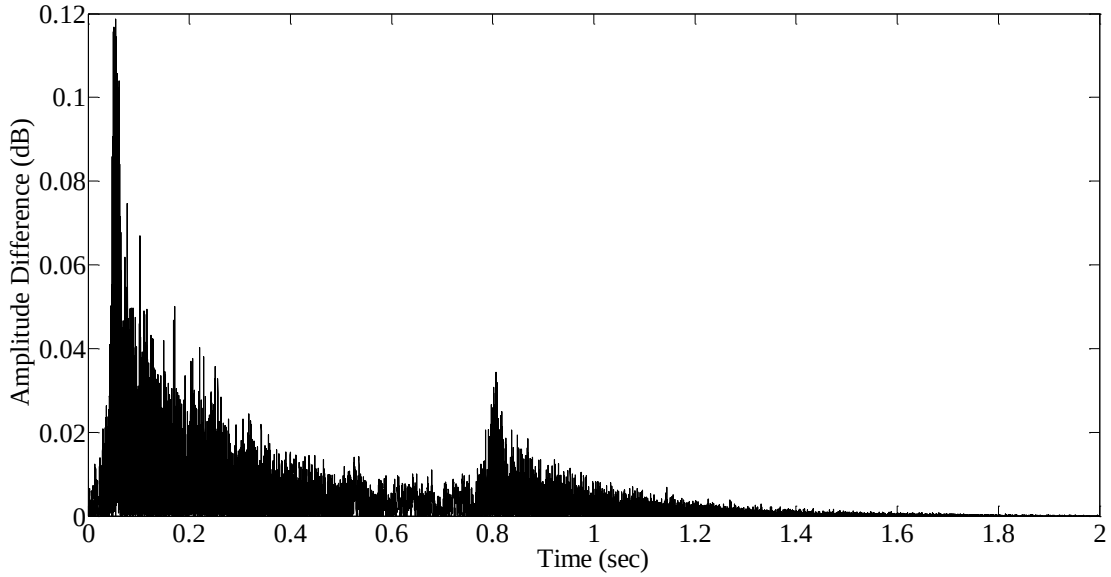


Figure 108 – Magnitude difference function between the reference RIR and the one measured using the stimulus presented in Figure 107.

C. Further Cases Study (Presto-chirps Method)

C1. Case Study: Broadcasting Auditorium “Sala A”

In this study, a medium-sized auditorium has been tested. It is rectangular-shaped with a high ceiling. Figure 109 shows a photo taken from the stage (at the loudspeaker’s position), during the measurements session. The auditorium is used for live broadcasting programs, which include the presence of an audience (up to 150 people, included the technical staff and the musicians). The stage is about 120 square meters and it is used to host live bands. The room is fitted with 95 upholstered seats. The rooms are spotted by sound diffractive elements. The floor is covered by a plaster-like reflective surface. Figure 110 shows its architectural drawing, with the positions of the loudspeaker and the microphone.

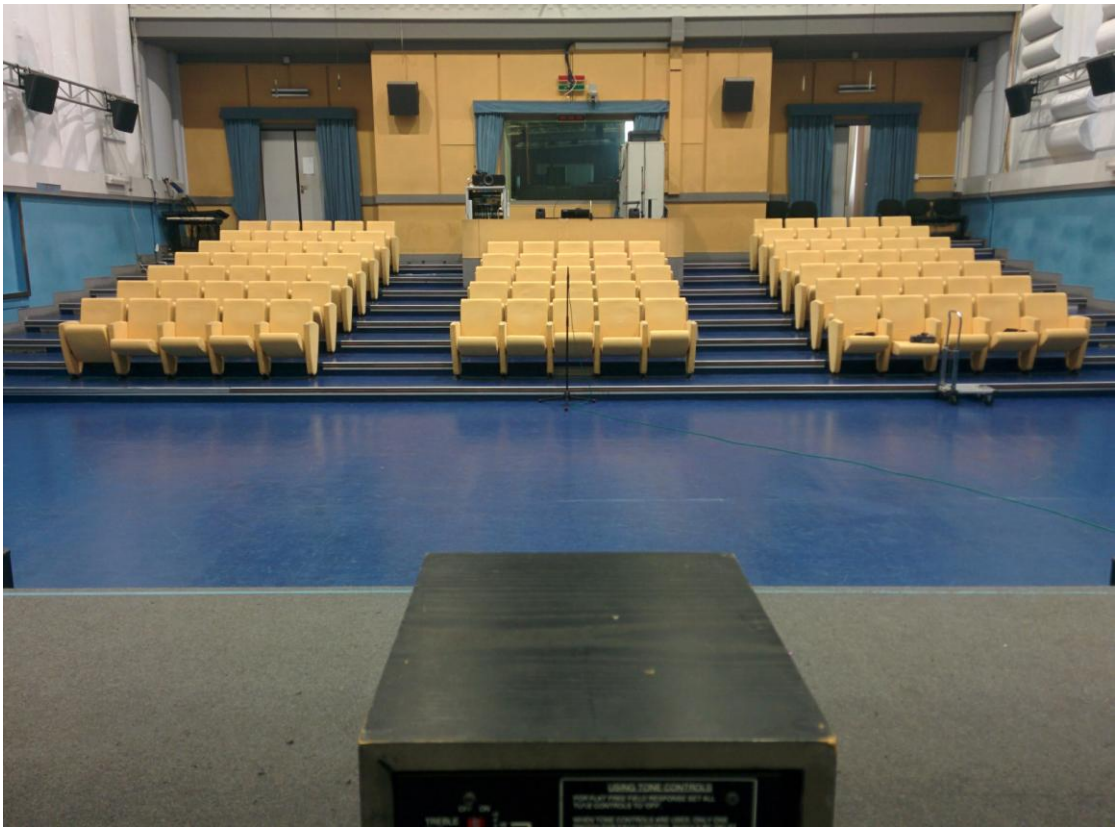


Figure 109 – Photo of the tested auditorium.

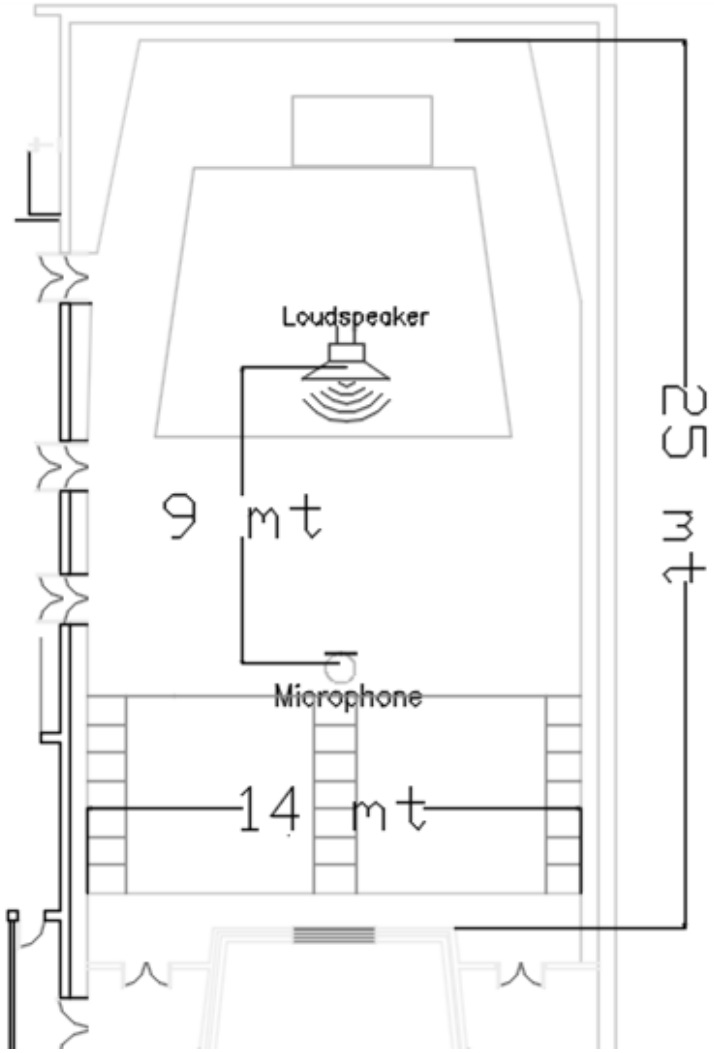


Figure 110 - Architectural drawing of the tested room.

Figure 111 shows the reference RIR and the FRF, in upper and lower panels, respectively. The measured reverberation time RT30, averaged within the sub bands from 63 Hz to 4 kHz, was approximately 1.07 seconds. The presto-chirps method used the musical stimulus presented in Section 3.6.3, and an analysis window of one second.

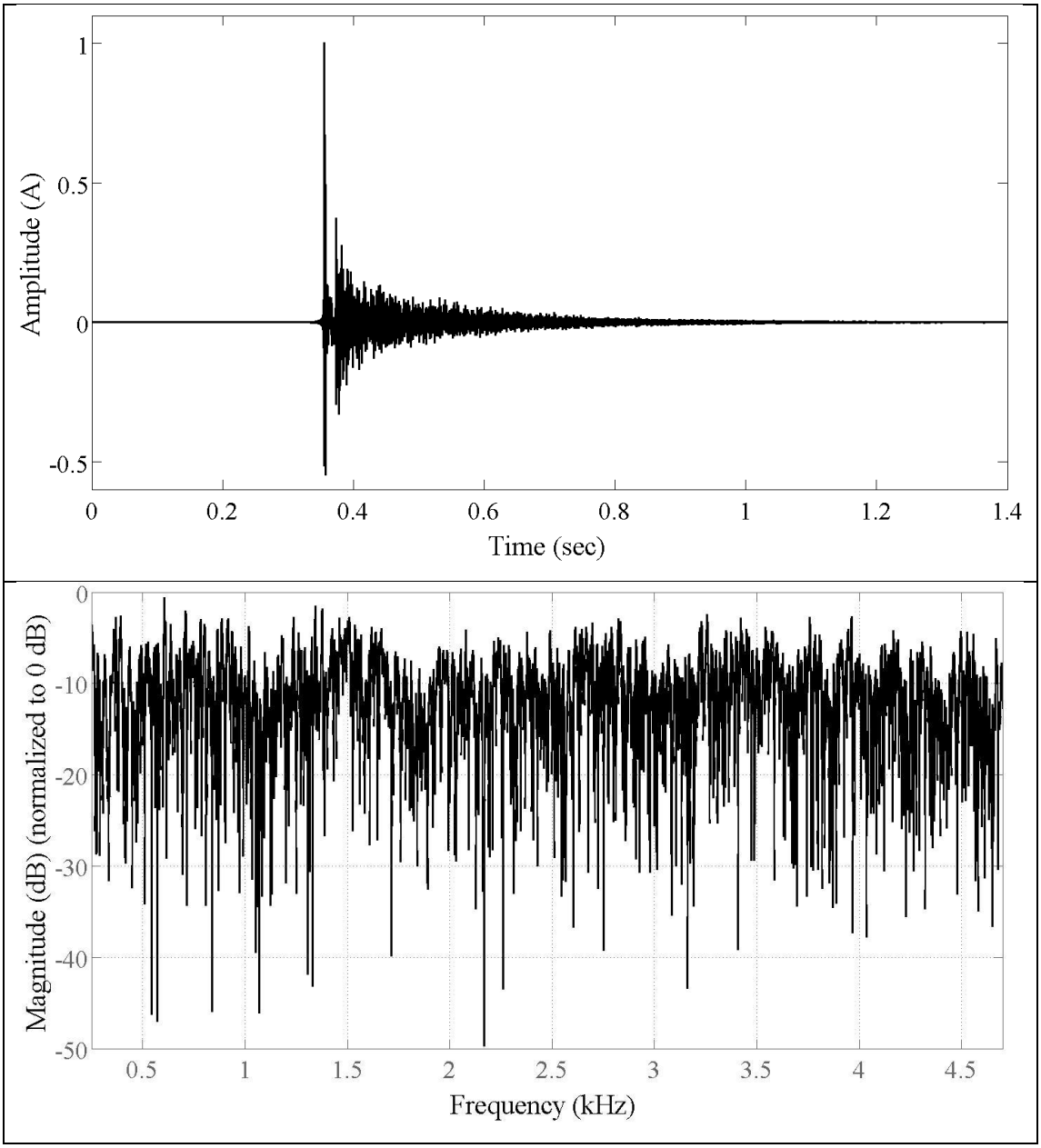


Figure 111 - Architectural drawing of the tested room.

Acoustic Parameters Analysis

Table 16 presents the comparison between the reference acoustic parameters and those measured using the presto-chirps method. The octave band from 63 Hz up to 4 kHz are reported. Similarly to the in situ measurement presented in Section 4.3, for α=1 the measured acoustic parameters show differences with the reference values within DLs. This is except for the parameters RT20 and RT30 in the 63 Hz sub bands,

where the SNR of the reference RIR was very low, as it can be seen in the EDCs shown in Figure 112. In the sub bands from 63 Hz to 250 Hz, for α≠1, the synthesis noise provokes an overestimation of all the tested acoustic parameters. Conversely, in the higher sub bands, the measured acoustic parameters are within the reference DLs.

C80 [dB] *(DL = 1 dB)*							
Bands [Hz]	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
Reference	4.52	7.96	5.62	5.12	4.95	7.15	5.94
α = 1	0.22	0.22	0.29	0.15	0.49	0.23	0.06
α = 0.75	0.47	1.66	0.13	0.13	0.48	0.34	0.17
α = 0.5	1.34	2.69	0.33	0.18	0.88	0.63	0.28
α = 0.25	1.38	3.22	1.02	0.45	0.40	0.42	0.16
α = 0	1.83	3.11	1.83	0.59	0.34	0.31	0.12

D50 [%] *(DL = 5%)*							
Bands [Hz]	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
Reference	60.34	71.37	70.31	60.42	63.92	77.81	72.04
DL 5% [%]	3.37	3.02	3.57	3.52	3.02	3.20	3.89
α = 1	3.11	0.21	1.91	1.73	3.01	1.06	0.10
α = 0.75	4.14	3.37	0.42	1.85	3.20	1.48	0.78
α = 0.5	8.60	6.31	1.39	2.54	6.03	2.59	1.11
α = 0.25	7.69	8.31	5.06	1.21	4.19	1.46	0.81
α = 0	10.97	8.02	3.83	0.97	4.41	1.19	0.86

EDT [sec] *(DL = 3.2% ± 0.6%)*							
Bands [Hz]	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
Reference	1.19	0.783	0.983	0.82	0.946	0.883	1.066
DL 3% [sec]	0.036	0.023	0.029	0.025	0.028	0.026	0.032
α = 1	0.030	0.024	0.002	0.002	0.015	0.001	0.012
α = 0.75	0.022	0.278	0.005	0.006	0.016	0.014	0.017
α = 0.5	0.071	0.425	0.026	0.017	0.027	0.022	0.033
α = 0.25	0.195	0.495	0.062	0.035	0.026	0.025	0.022
α = 0	0.195	0.489	0.067	0.064	0.052	0.024	0.016

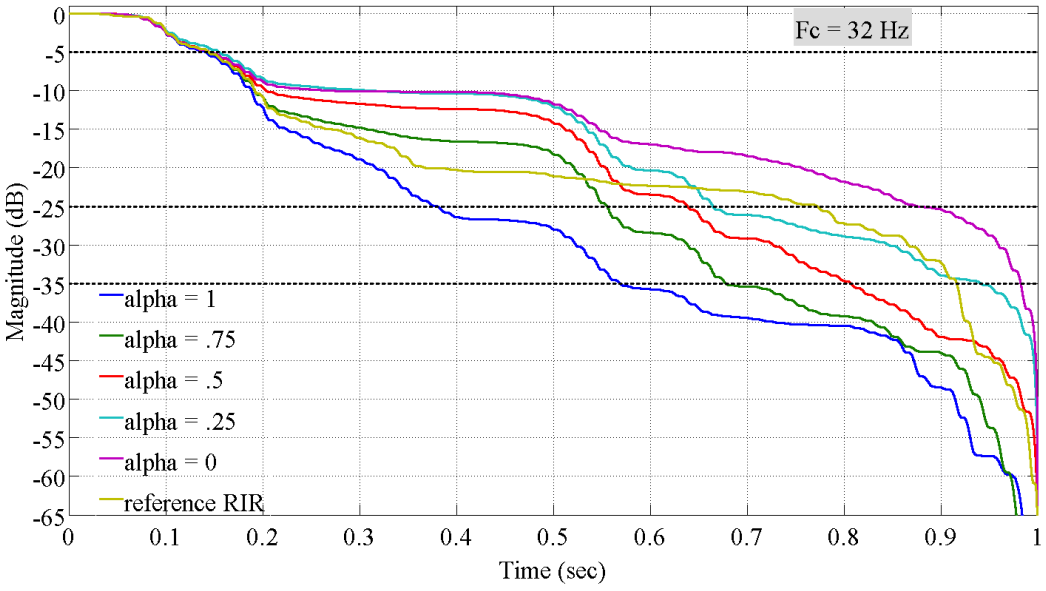
RT20 [sec] *(DL = 3.2% ± 0.6%)*							
Bands [Hz]	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k***
Reference	0.71	0.81	0.74	0.91	1.03	1.20	1.23
DL 3% [sec]	0.021	0.024	0.022	0.027	0.031	0.036	0.037
α = 1	0.082	0.001	0.01	0.009	0.006	0.016	0.005
α = 0.75	0.359	0.001	0.002	0.017	0.007	0.025	0.007
α = 0.5	0.577	0.038	0.016	0.021	0.019	0.02	0.007
α = 0.25	0.679	0.075	0.054	0.028	0.02	0.012	0.011
α = 0	0.497	0.259	0.065	0.012	0.013	0.018	0.009

RT30 [sec] (DL = 3.2% ± 0.6%)	Bands [Hz]	63	125	250	500	1k	2k	4k*
	Reference	0.92	0.88	0.78	0.93	1.06	1.21	1.22
	DL 3% [sec]	*0.028*	*0.026*	*0.023*	*0.028*	*0.032*	*0.036*	*0.037*
	α = 1	0.138	0.012	0.005	0.027	0.009	0.002	0.007
	α = 0.75	0.114	0.013	0.018	0.029	0.016	0.002	0.011
	α = 0.5	0.247	0.034	0.007	0.043	0.022	0.001	0.015
	α = 0.25	0.285	0.121	0.053	0.05	0.028	0.009	0.003
	α = 0	0.112	0.172	0.069	0.003	0.024	0.009	0.003

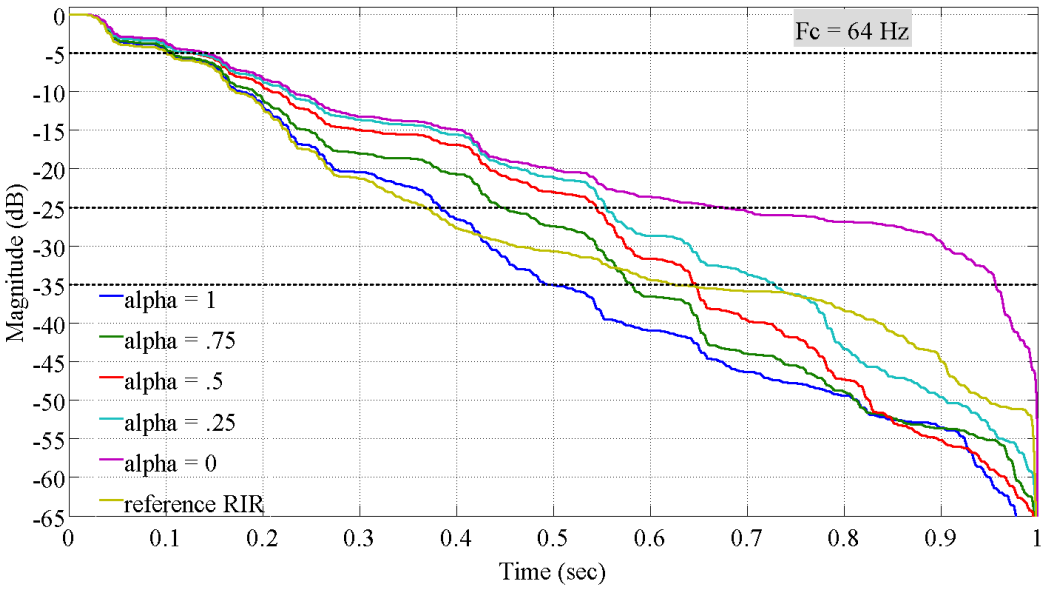
Table 16 – Acoustic parameters estimation in the octave bands from 63 Hz to 2 kHz, and in the third octave band centred at 4 kHz - for several α values.

Figure 112 shows the EDCs of the reference and measured RIRs, in the sub bands from 31.5 Hz up to 4 kHz. In the 31.5 Hz and 63 Hz sub bands the reference RIR is strongly affected by the background noise, which makes the estimation of acoustical parameters unreliable.

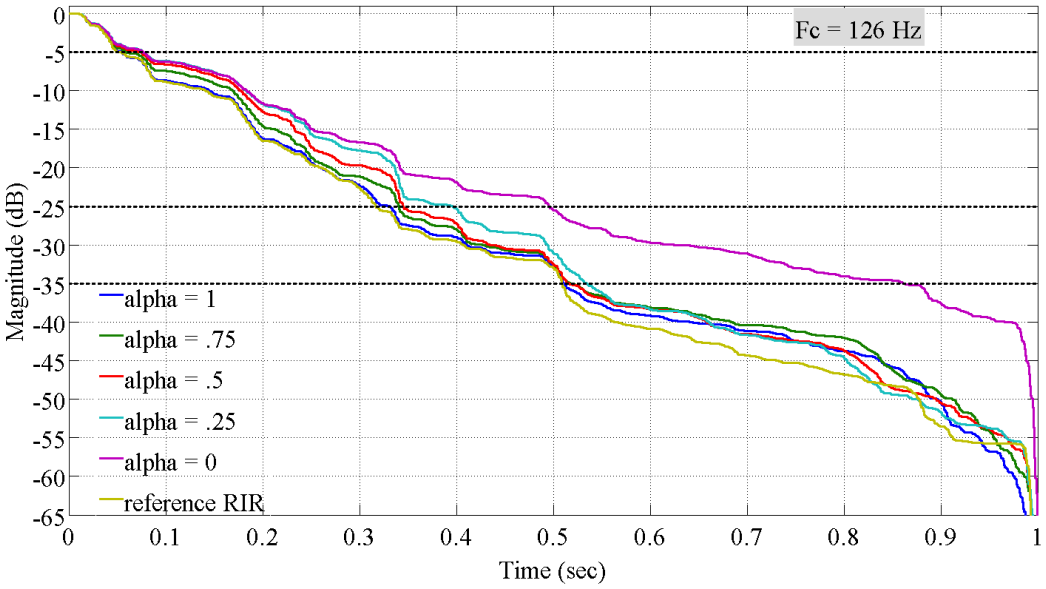
31.5 Hz



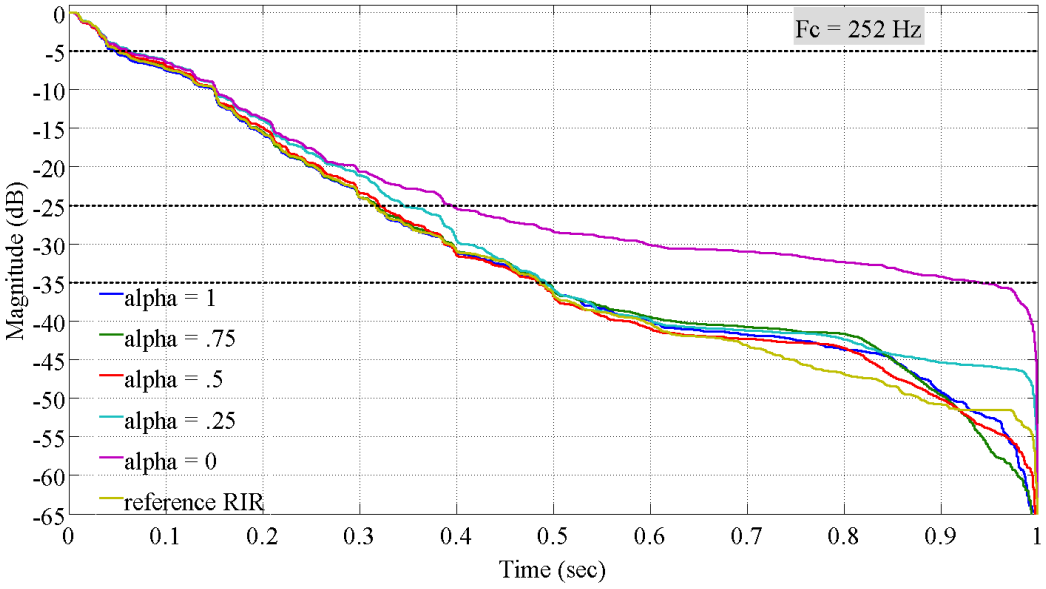
63 Hz



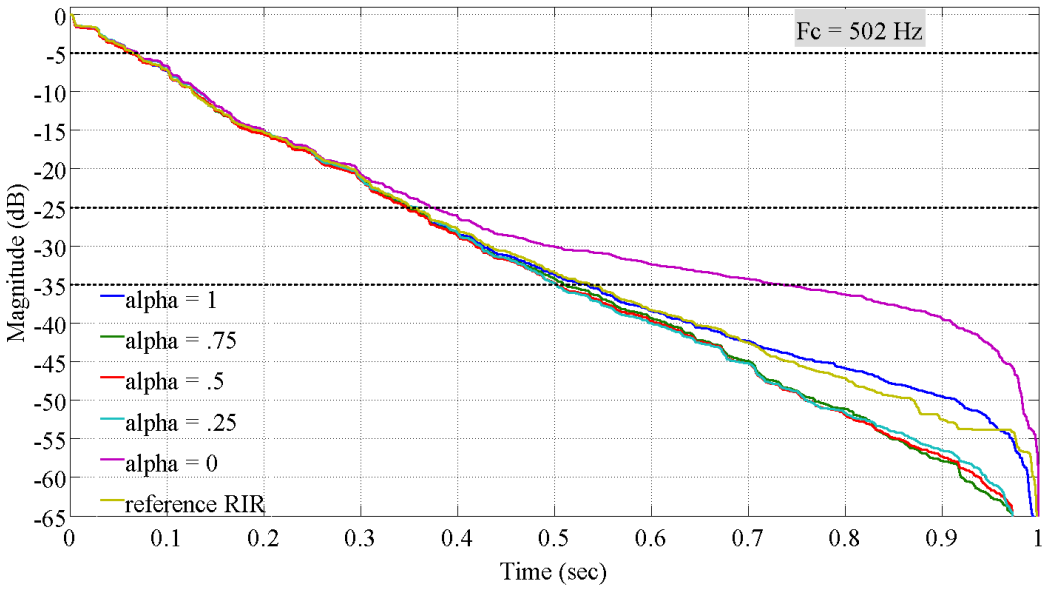
125 Hz



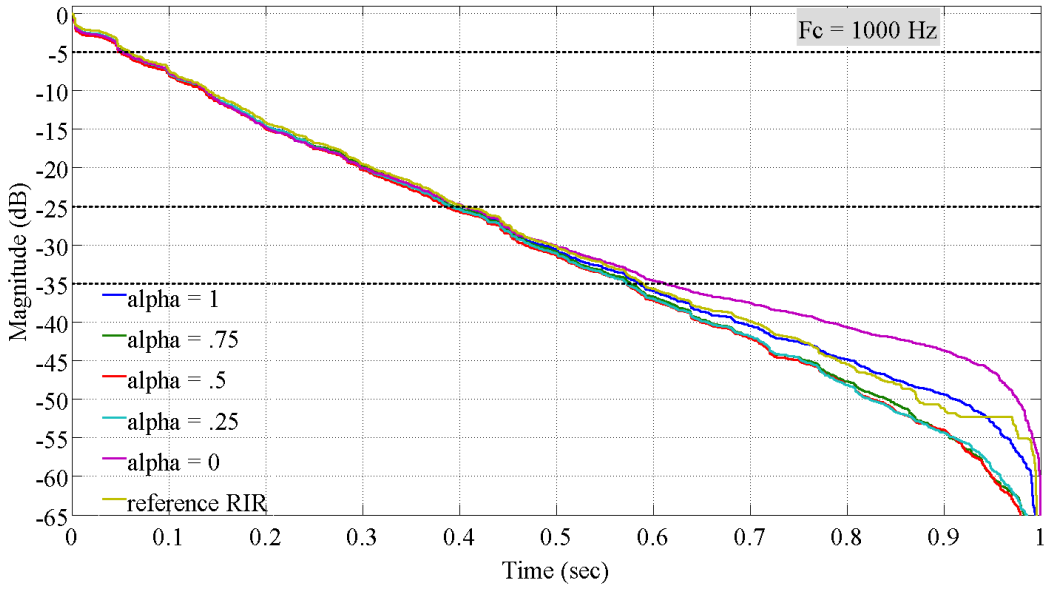
250 Hz



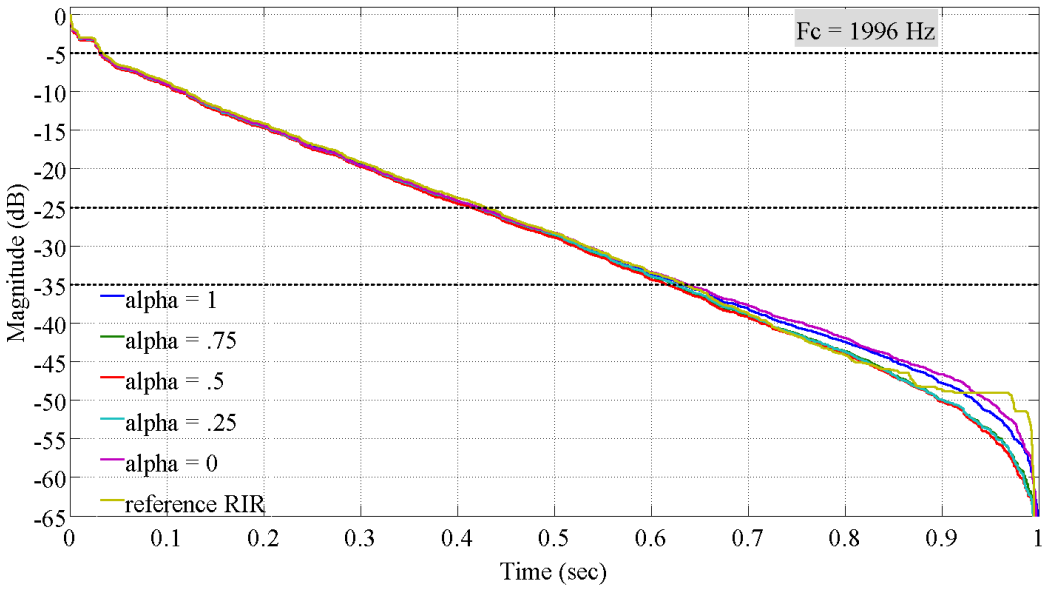
500 Hz



1 kHz



2 kHz



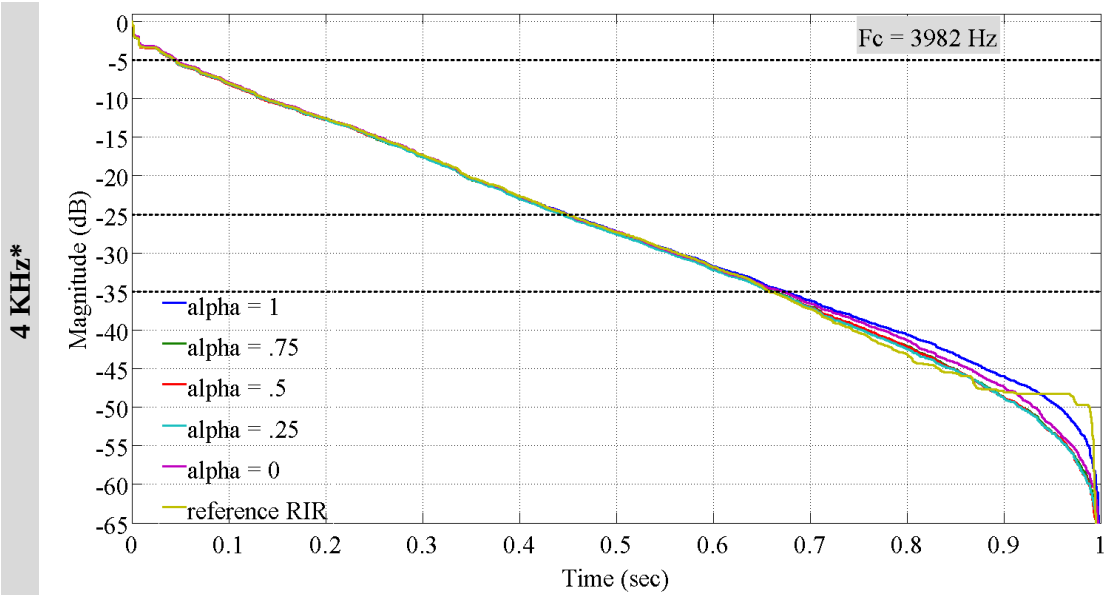


Figure 112 – EDCs of the measured RIRs (as function of α) plus the reference RIR. Each figure accounts for a different sub band as labelled in the left. The dashed lines indicate the decay range generally used to calculate the acoustic parameter RT30. The label 'Fc' represents the central frequency of the octave band filter used in each sub band.

Table 17 reports the SNRs for each tested stimuli and the reference RIR. The highest SNR value, of each sub band, is highlighted. It can be seen that for α=1, in the sub bands 31.5 Hz and 63 Hz the SNR is higher than the reference RIR, which further proves the averaging effect at those bands.

On the other hand, it can also be seen that the SNRs in the other sub bands show higher values for lower α. This trend, which has not been recorded in the other tested in situ measurements, might have been caused by the increasing narrowing of the transition bands of presto-chirps as α decreases. Narrow transition bands have higher noise rejection property. Further works are needed to shed light on this issue.

SNR [dB] (Not calibrated)	Bands [Hz]	31.5	63	125	250	500	1k	2k	4k*
	LIN chirp	30.6	39.3	42.6	39.0	41.6	40.1	37.2	36.3
	α = 1	39.8	47.4	43.6	39.7	39.5	39.6	38.4	36.4
	α = 0.75	37.9	41.8	41.4	43.7	49.5	45.5	40.7	39.7
	α = 0.5	30.4	41.3	41.4	41.8	48.0	45.8	41.2	39.8
	α = 0.25	21.3	40.0	41.2	32.6	46.6	45.2	40.9	39.9
	α = 0	16.8	20.8	27.3	22.2	29.6	33.5	37.3	38.8

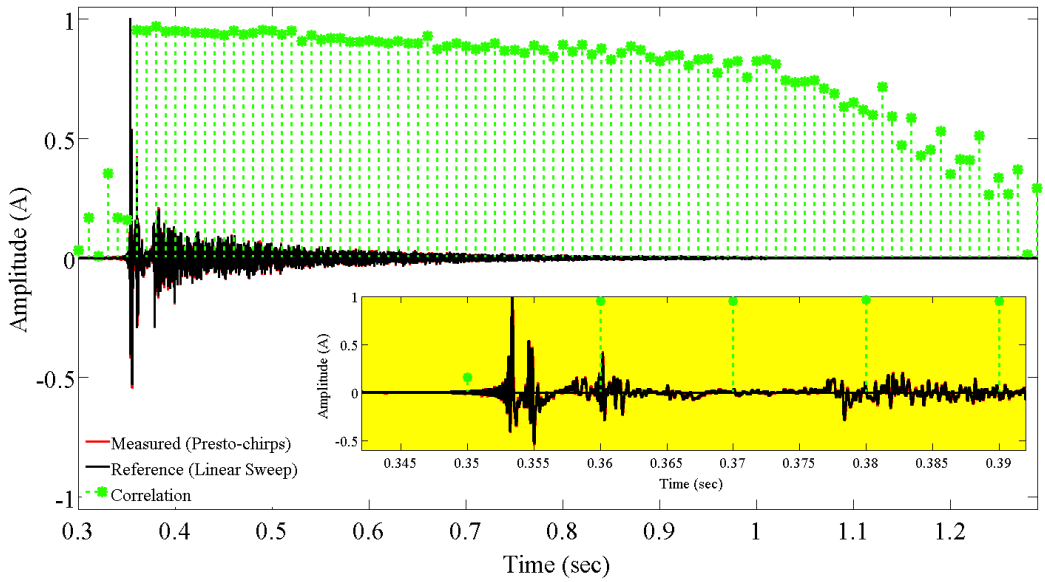
Table 17 – SNR given by the software Audacity. The first row, in blue, refers to the reference RIR measured using a 10 seconds long chirp (27 – 4.7 kHz).

Time Domain Analysis

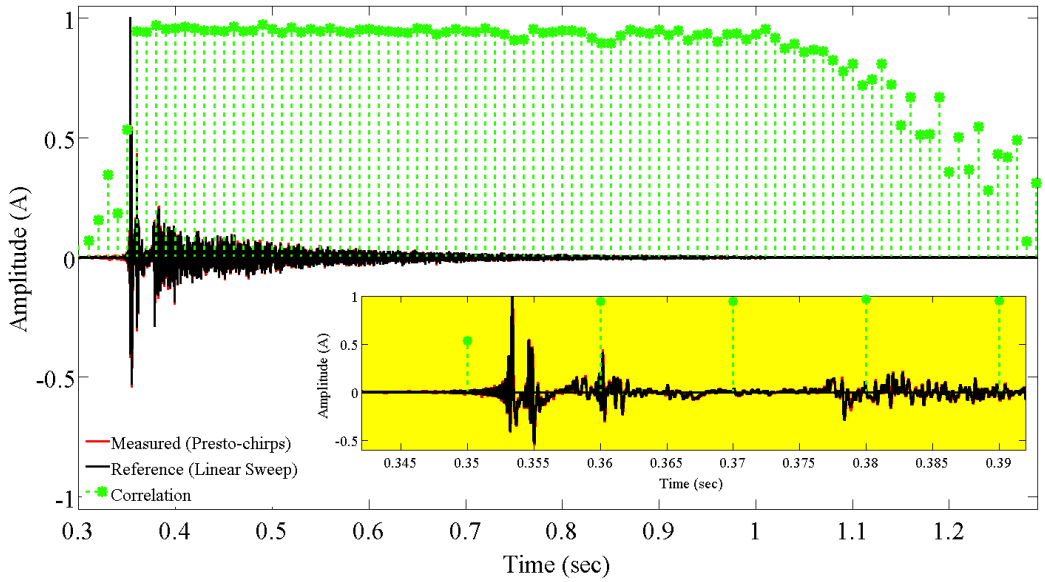
For α=1 the early part of the measured RIR strongly correlates with the reference RIR. On the contrary, the late part of the RIR is affected by background noise, which consequently decreases the degree of correlation. Figure 113 shows the reference and measured RIRs. It can be seen that the reference RIR is the one affected the most, among all the measured RIRs, at low frequency bands.

The slightly monotone decrease of the correlation values in the case α=1, may be caused by some errors during the measurements, for example time variance effect. The HAVC was switched on during the tests. It is not present in the comparisons for α≠1, which on the other hand presents an increasing lack of correlation as α decreases due to the synthesis noise.

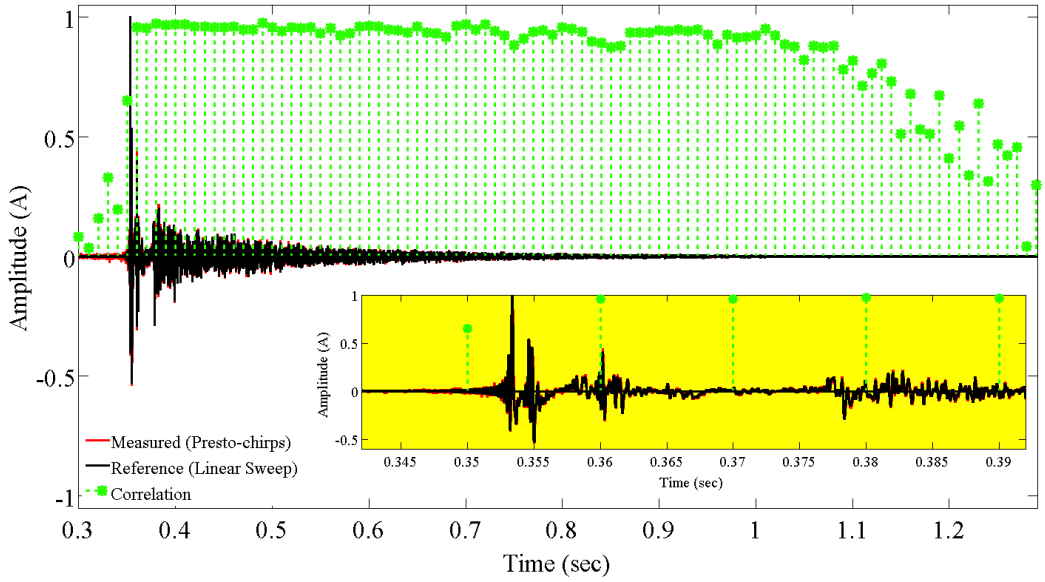
α = 1


Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

α = 0.75


Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

α = 0.5


Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

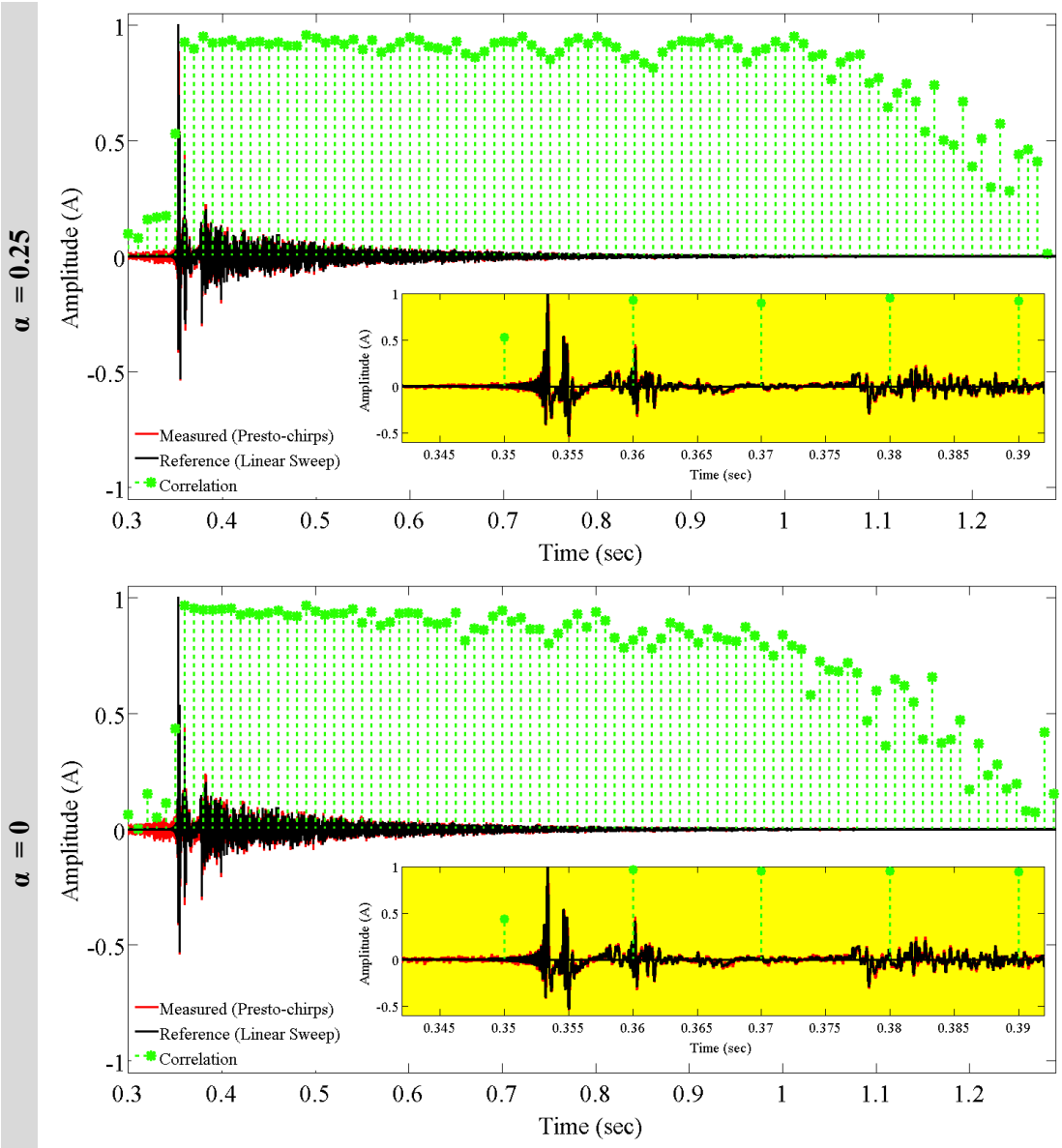


Figure 113 - Comparisons between the reference RIR (in black) and the measured RIRs (in red). The green marks point out the degree of correlation between 10 ms long blocks of the two RIRs. In the yellow boxes are shown the first 40 ms of the RIRs.

The correlations of the whole measured RIRs with the reference one are shown in Figure 114, versus the different Tukey windows used. As expected, the correlation is higher for α=1 compared with the other windows, which then shows a progressively decrease for lower α values.

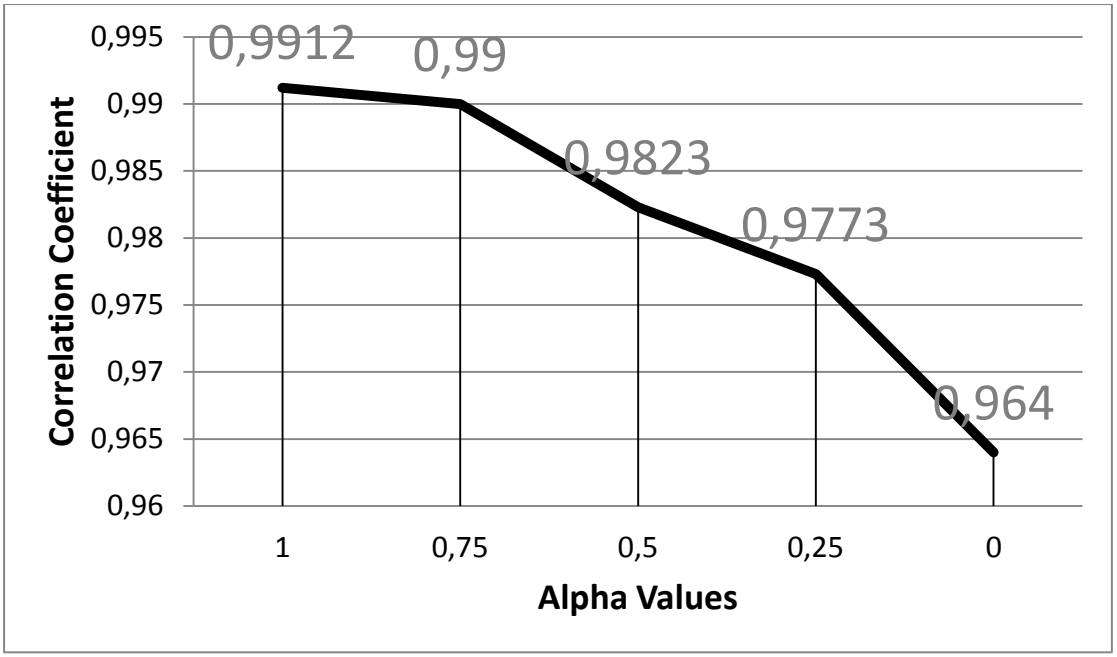
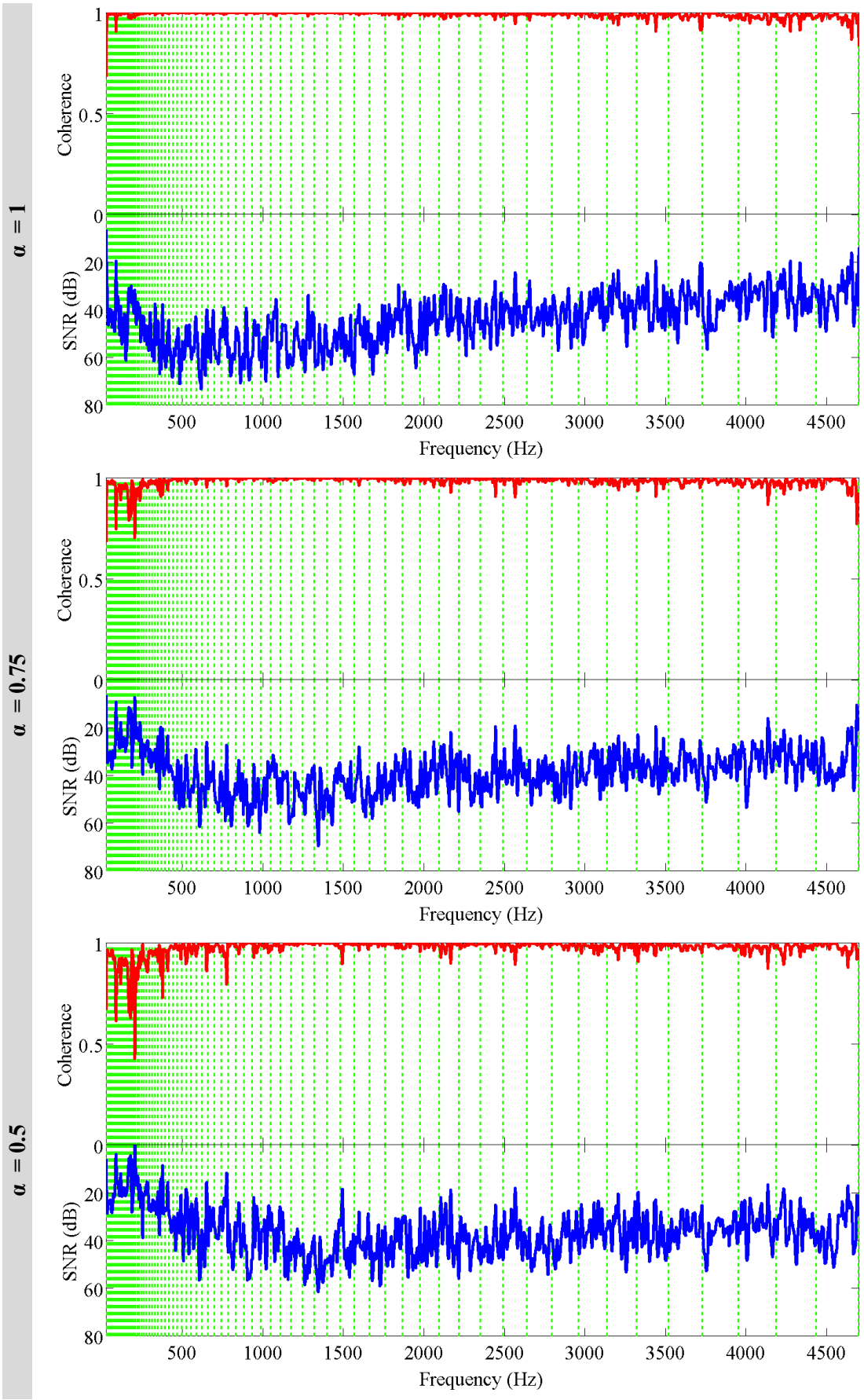


Figure 114 – Degree of correlation between the reference RIR and the RIRs measured with the different α-windowed stimuli.

Frequency Domain Analysis

Figure 115 shows the coherence and SNR analyses. The Best result is obtained from the use of Hanning-windowed presto-chirps (α=1). The SNR at the high end of the measured spectrum was not too high, which consequently caused low coherence. As α decreases, the coherence and the SNR lowers progressively, starting from low frequencies. The frequencies showing minor coherence and SNR are related to the frequencies of the reference FRF having lower magnitude (bottom panel of Figure 111).


α = 1
α = 0.75
α = 0.5
Coherence
SNR (dB)
Frequency (Hz)
1
0.5
0
20
40
60
80
500
1000
1500
2000
2500
3000
3500
4000
4500

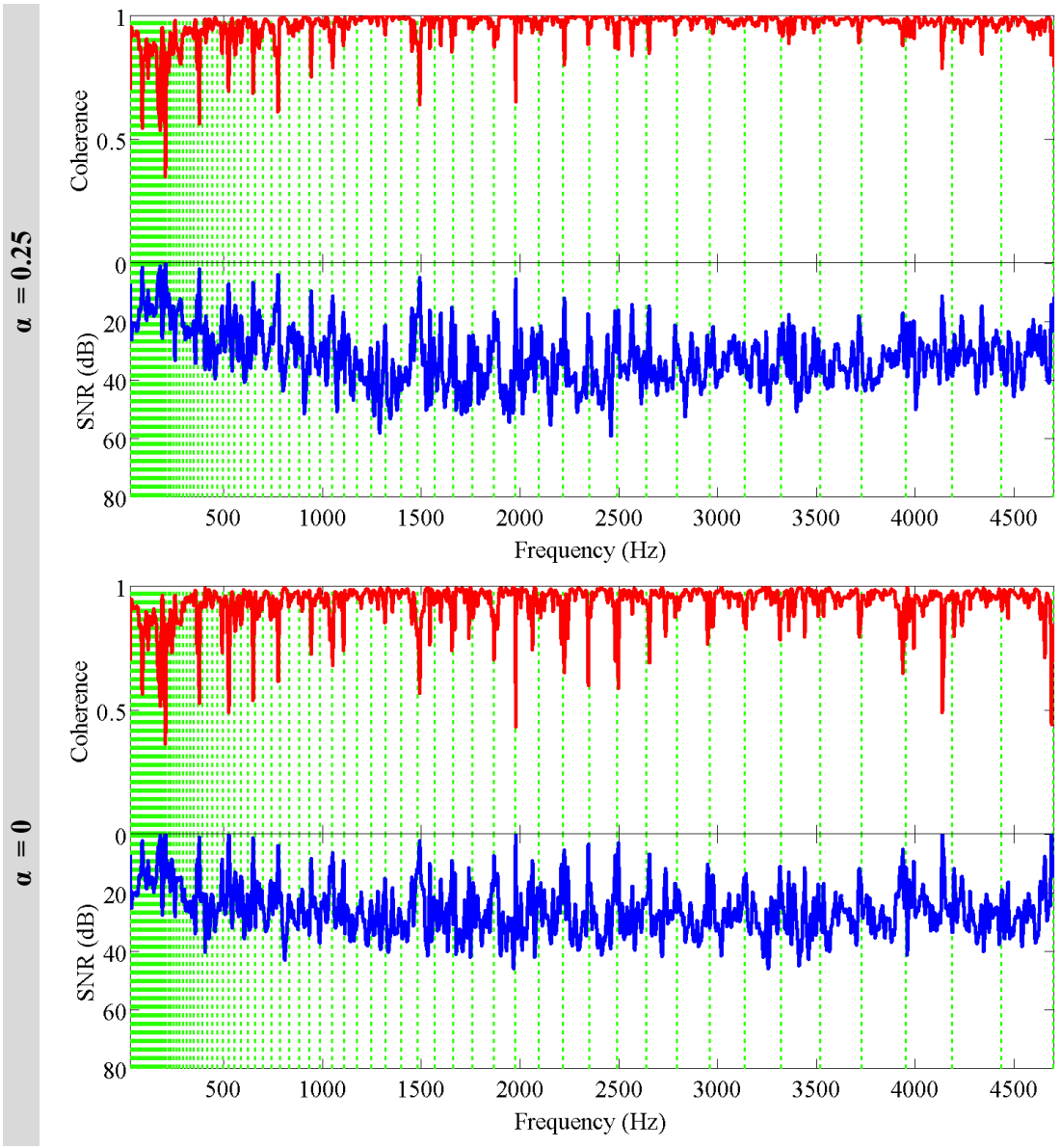


Figure 115 – In red is shown the coherence function, and in blue the SNR, between the reference FRF and the FRF calculated using different α related stimuli, as labelled.

Discussions of the Case Study

The acoustic analysis of the auditorium has confirmed that Hanning-window presto-chirps give the most accurate measurement of the RIR. For the other window configurations, bias is increasingly introduced both in time and in frequency domain as α decreases.

Presto-chirps stimuli (mainly the Hanning-windowed) can obtain, at the frequency bands 31.5 Hz, 63 Hz, and 125 Hz, a SNR about 10 dB higher than the one obtainable using a 10 second-long linear chirp. However, it has also been highlighted that the SNR in the frequencies bands above 250 Hz is higher when sharper window configurations are used (about 3 dB higher than the both the reference RIR and the RIR measured using α=1 presto-chirps). This effect may be related to the narrower transition bands of presto-chirps for lower α, which may emphasise the rejection of the off-band noise.

In conclusion, for the tested case, Hanning-windowed presto-chirps have resulted superior to the other Tukey-windowed presto-chirps.

C2. Case Study: Studio Room “Sala U3”

This case study presents the measurements of the studio room in a broadcasting radio company. The room is used for live radio programs, which includes the presence of a small audience (about 20 people), plus musicians, presenters and guests. The room is box-shaped, with a 3 meters high ceiling. The space is filled with office-chairs, musical instruments (included a drum booth), loudspeakers and an elliptical table surrounded by several office-style chairs. Figure 116 shows a photo of the room, and Figure 117 its architectural drawing, with the reported dimensions of the room and the distance between the loudspeaker and the microphone.

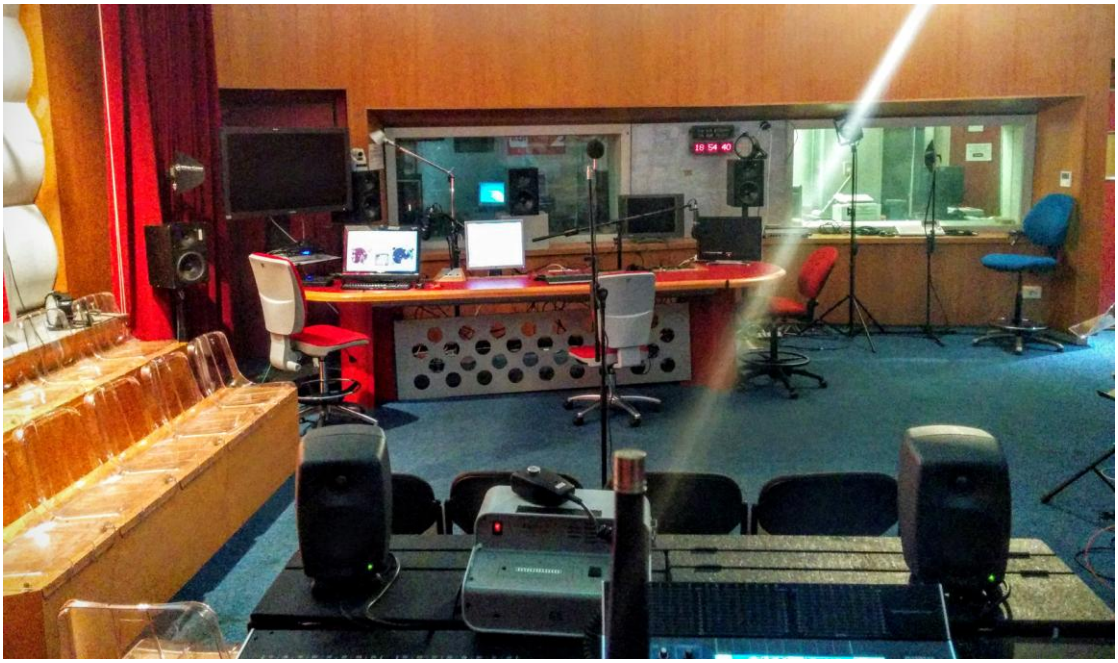


Figure 116 – Photograph of the tested room, taken before the measurement session.

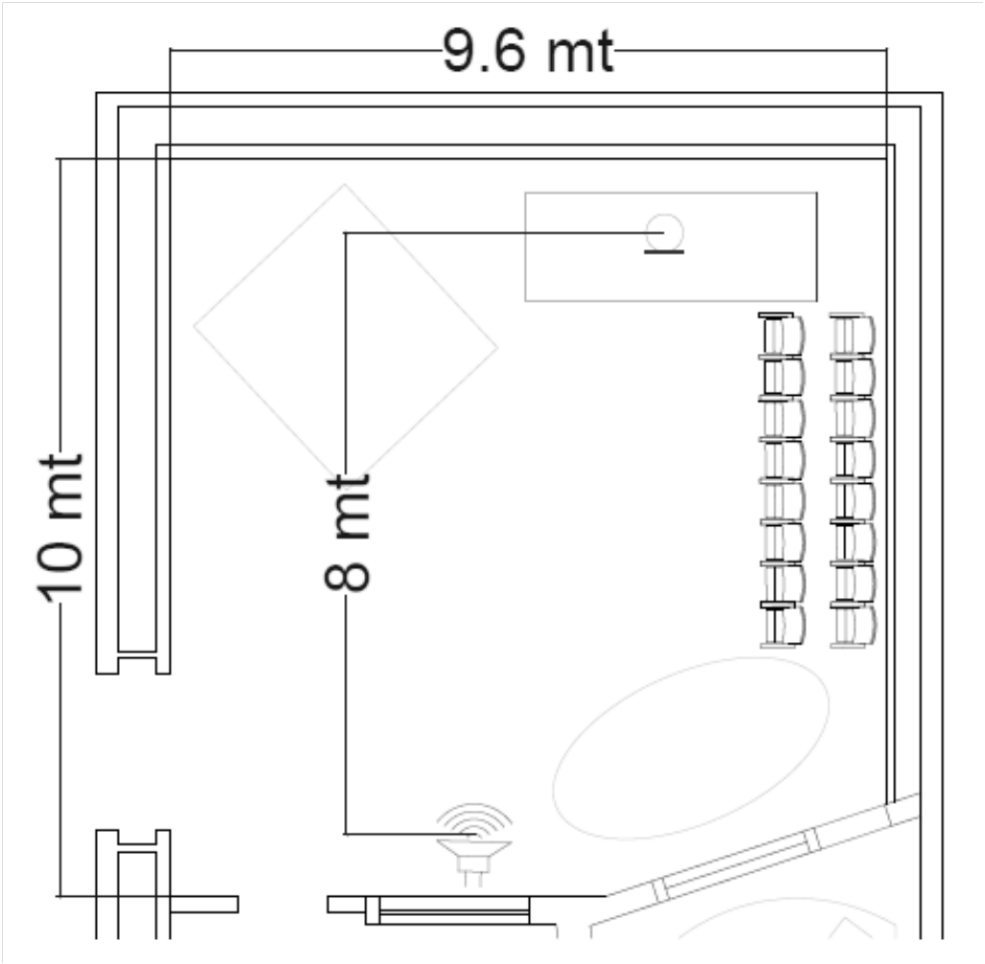


Figure 117 - Architectural drawing of the tested room.

The measured reference RT30 (averaged within the octave bands from 63 Hz up to 4 kHz) resulted to be 0.43 sec. Figure 118 reports, in the upper panel, the impulse response of the room; in the bottom panel its frequency response. An analysis window of 0.s seconds has been used.

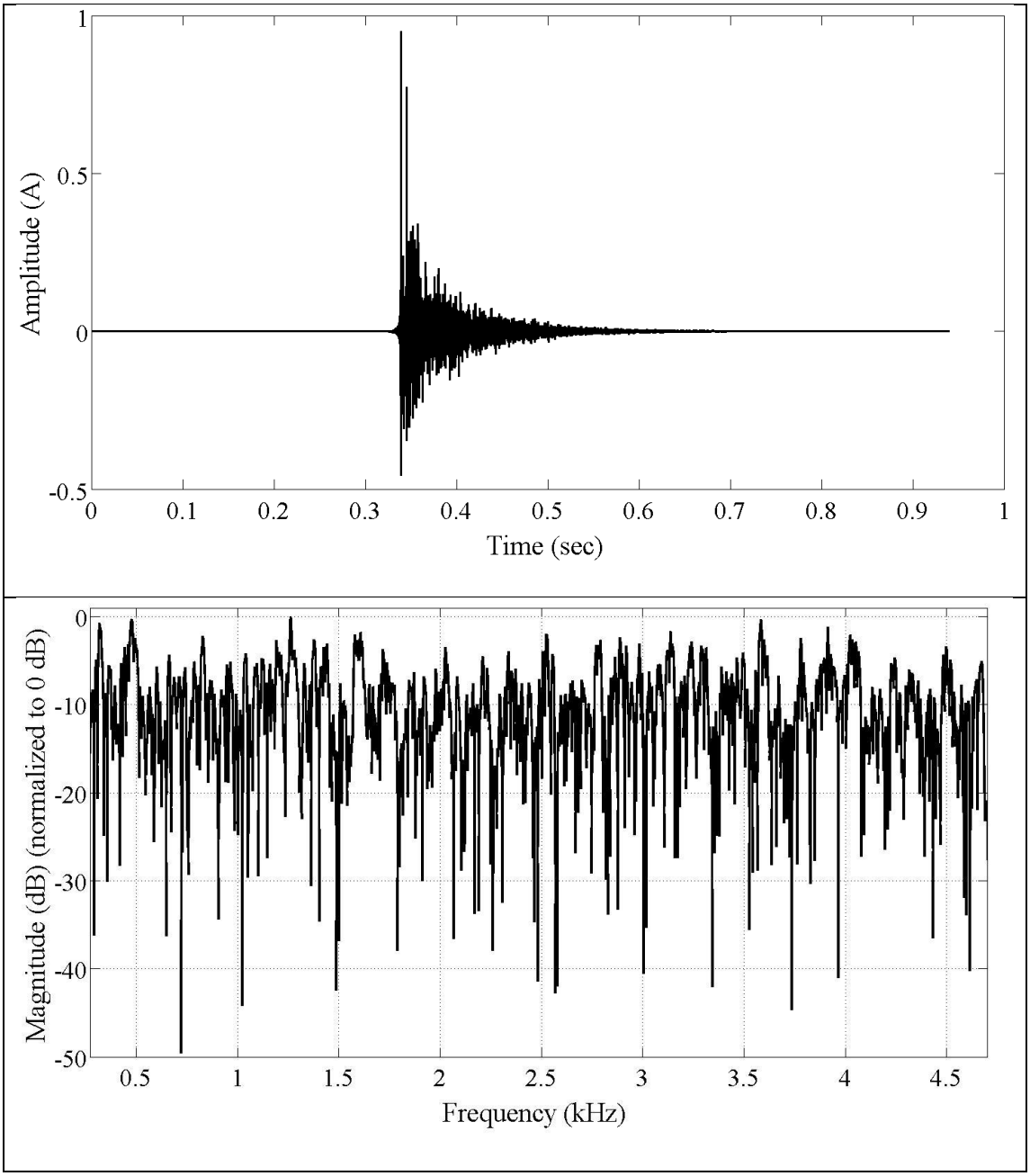


Figure 118 – The upper and bottom panels shown the impulse and frequency responses of the tested room, respectively.

Acoustic Parameters Analysis

With the exception of the sub band centred at 63 Hz, where the SNR of the reference RIR was too low, only α=1 the measured acoustic parameters show differences with reference values smaller that the DLs. For α≠1 only the sub bands from 500 Hz to 4 kHz, the measured acoustic parameters have difference smaller than the DLs with the reference values.

C80 [dB] *(DL = 1 dB)*

Bands [Hz]	63	125	250	500	1k	2k	4k*
Reference	13.052	11.67	16.548	13.311	13.734	12.053	12.179
α = 1	2.137	0.074	0.564	0.582	0.067	0.378	0.056
α = 0.75	2.447	0.729	0.582	0.557	0.23	0.036	0.085
α = 0.5	2.738	1.611	5.61	0.234	0.19	0.19	0.141
α = 0.25	2.896	2.375	6.647	0.516	0.303	0.007	0.295
α = 0	2.758	2.816	7.051	0.823	0.715	0.391	0.086

D50 [%] *(DL = 5%)*

Bands [Hz]	63	125	250	500	1k	2k	4k*
Reference	84.738	87.313	88.342	86.107	85.88	87.52	85.366
DL [%]	*4.237*	*4.366*	*4.417*	*4.305*	*4.294*	*4.376*	*4.268*
α = 1	4.111	0.694	3.672	0.331	0.686	1.331	0.883
α = 0.75	4.531	1.553	9.43	0.554	0.567	0.191	0.448
α = 0.5	4.788	2.924	14.028	0.084	0.589	0.637	0.272
α = 0.25	5.17	4.448	16.406	1.062	1.58	0.973	0.293
α = 0	4.278	5.069	16.066	2.1	3.291	1.856	0.542

EDT [sec] *(DL = 0.024 sec)*

Bands [Hz]	63	125	250	500	1k	2k	4k*
Reference	0.462	0.344	0.358	0.408	0.365	0.382	0.391
α = 1	0.047	0.006	0.019	0.002	0.001	0.017	0.016
α = 0.75	0.049	0.037	0.088	0.002	0.003	0.016	0.016
α = 0.5	0.055	0.077	0.133	0.011	0.014	0.019	0.006
α = 0.25	0.061	0.168	0.133	0.017	0.022	0.017	0.002
α = 0	0.052	0.213	0.132	0.044	0.05	0.024	0.004

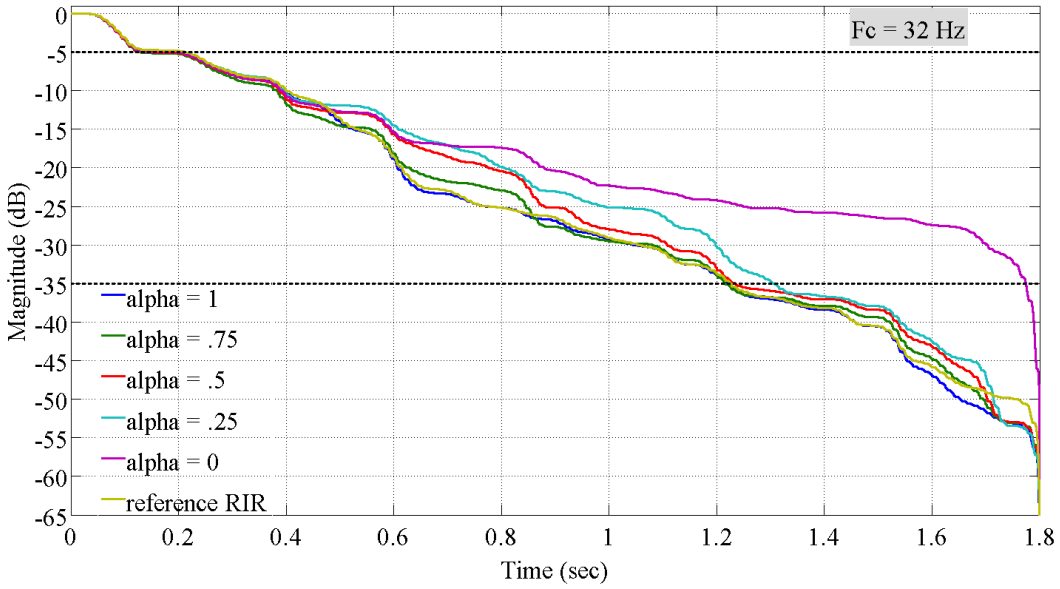
RT30 [sec] *(DL = 0.024 sec)*

Bands [Hz]	63	125	250	500	1k	2k	4k*
Reference	0.563	0.447	0.416	0.397	0.417	0.45	0.449
α = 1	0.054	0.023	0.024	0.015	0.009	0.01	0
α = 0.75	0.065	0.133	0.024	0.008	0.006	0.002	0
α = 0.5	0.092	0.155	0.033	0.007	0.002	0.005	0.002
α = 0.25	0.169	0.185	0.039	0.026	0.002	0.003	0.001
α = 0	0.197	0.177	0.042	0.03	0.006	0.015	0.004

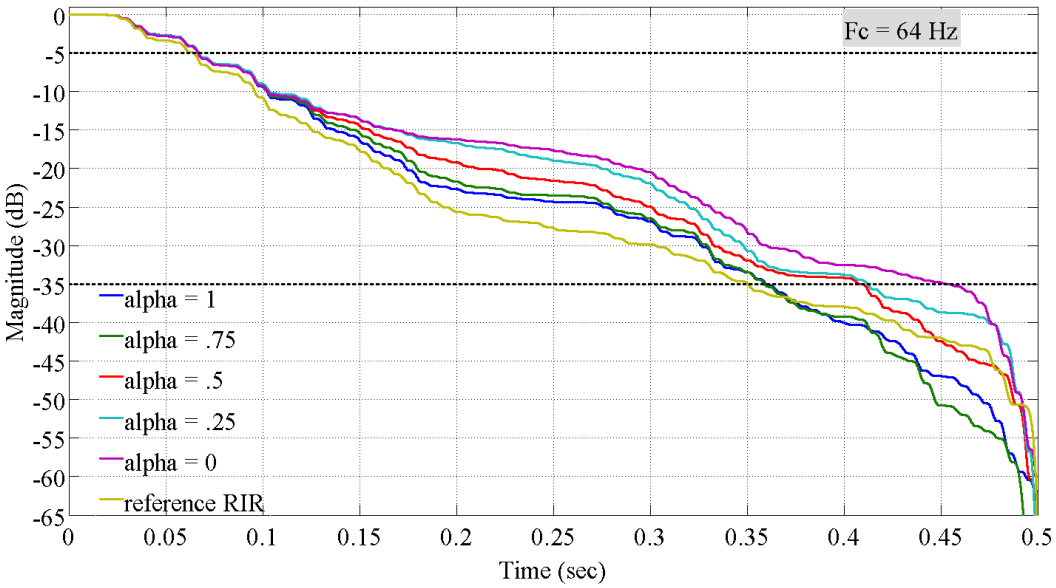
Table 18 – Acoustic parameters estimation in the octave bands from 63 Hz to 2 kHz, and in the third octave band centred at 4 kHz - for several α.

Figure 119 reports the EDCs calculated in the octave bands from 31.5 Hz to 4 kHz; each panel accounts for a specific sub band as labelled in the left. the octave band filter central frequencies (Fc) , are indicated in the grey box. The dashed horizontal lines, at -5 and -35 dB, indicate the decay range commonly used to calculate the RT30.

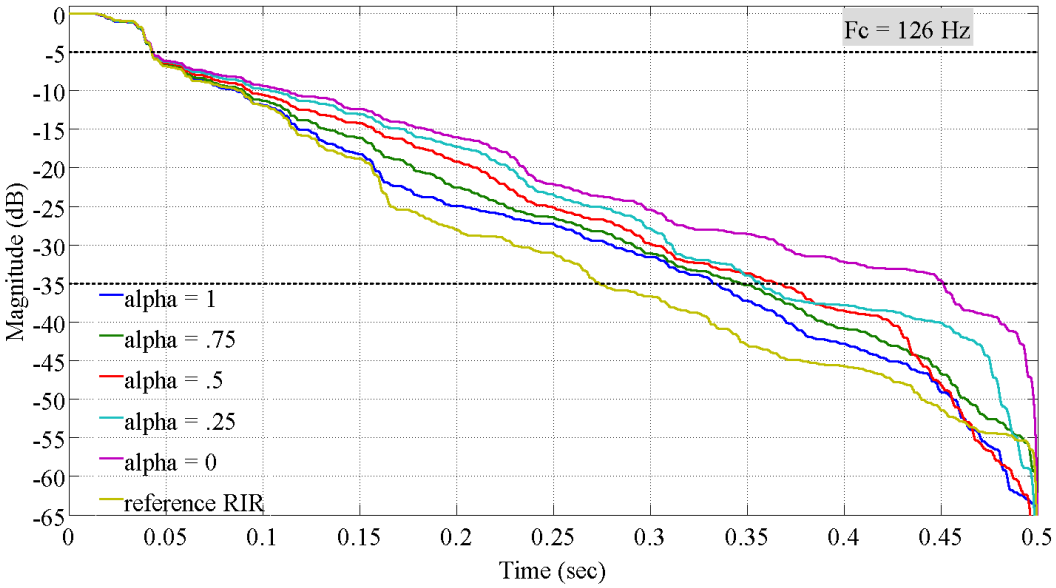
31.5 Hz



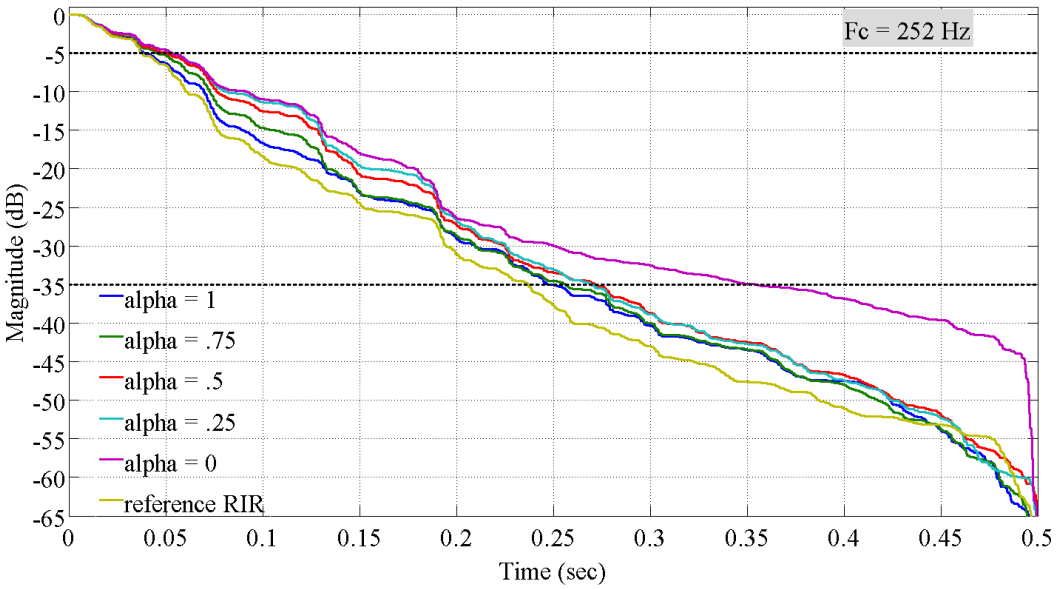
63 Hz



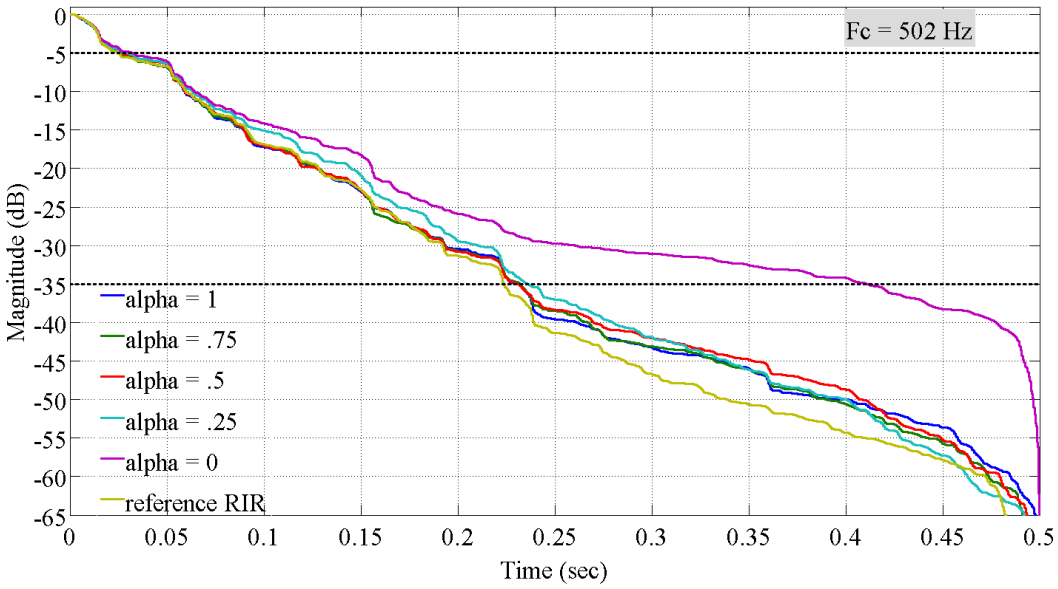
125 Hz



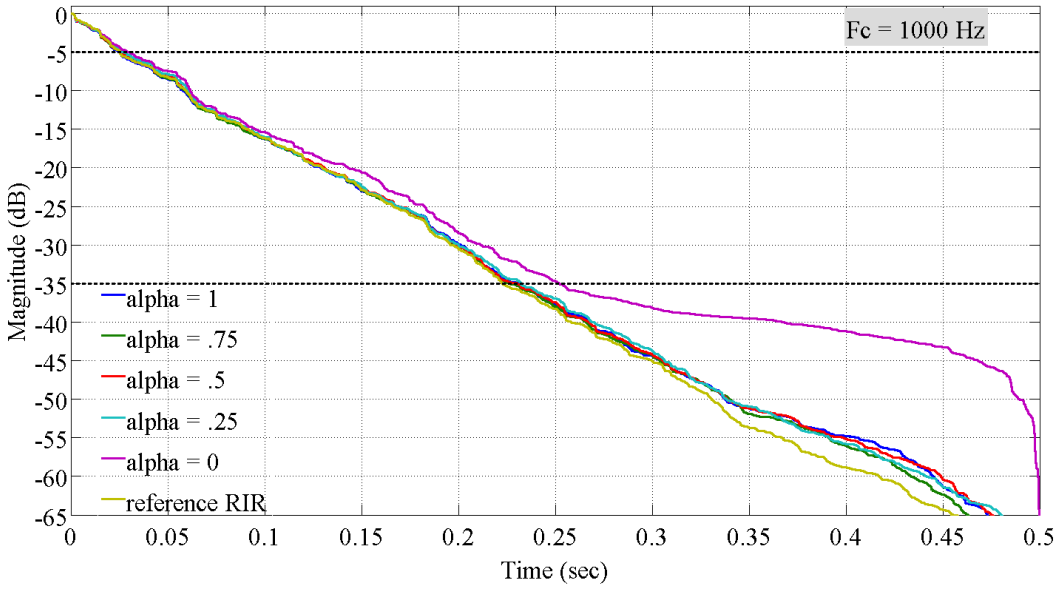
250 Hz



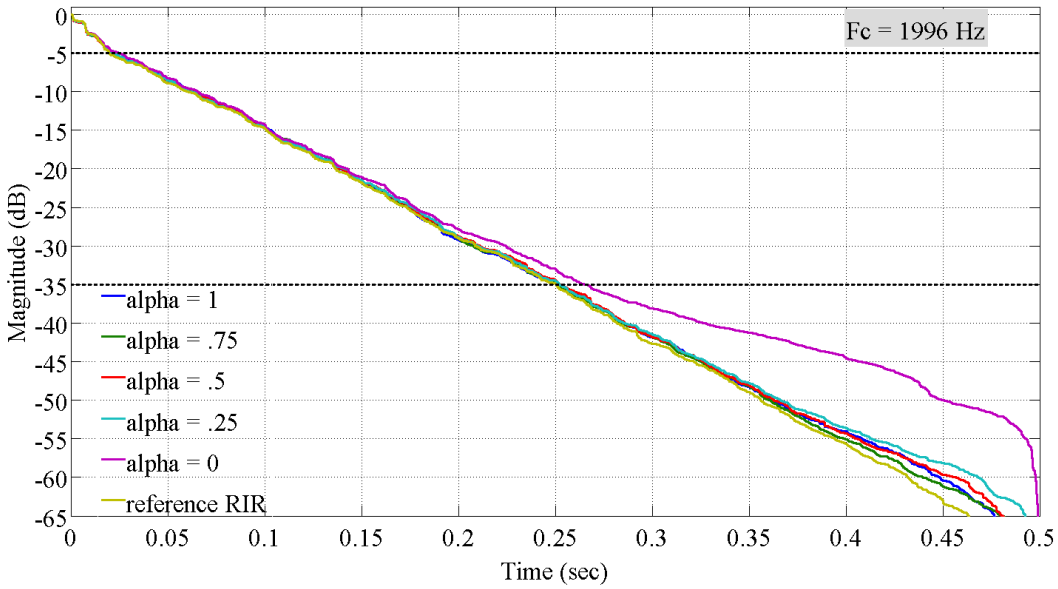
500 Hz



1 kHz



2 kHz



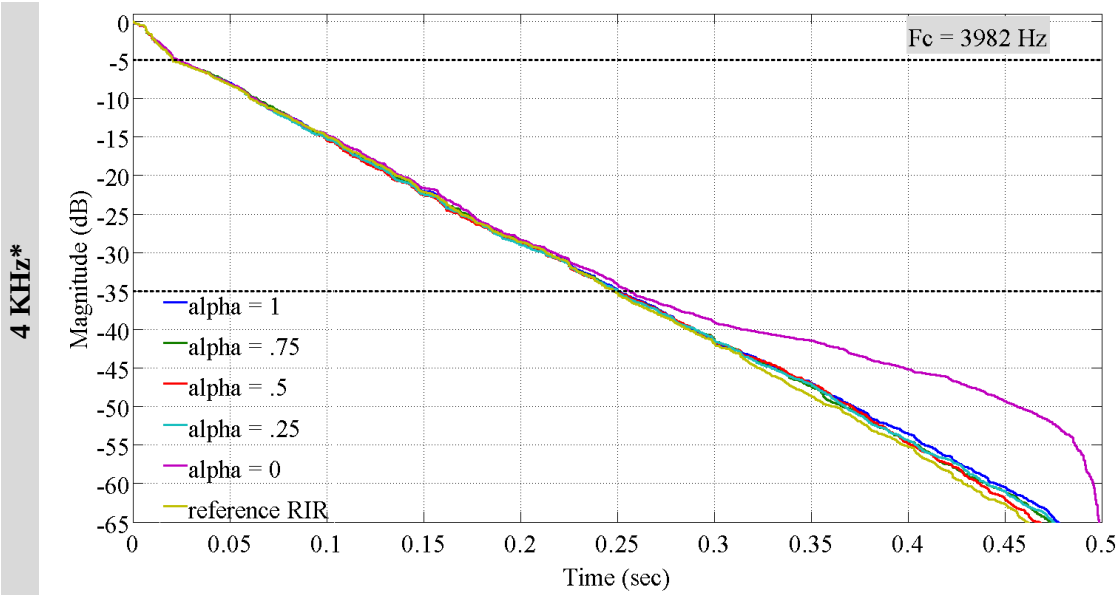


Figure 119 – EDCs of the measured RIRs (as function of α) plus the reference RIR. Each figure accounts for a different sub band as labelled in the left. The dashed lines indicate the decay range, which is commonly used to calculate the acoustic parameter RT30. The label 'Fc' represents the central frequency of the octave band filter used.

Table 19 reports the SNRs for each measured RIRs and the reference one. For α=1, in the sub bands 31.5 Hz and 63 Hz the SNR is higher than the reference RIR, which should be due to the averaging effect discussed in Section 4.2.1.5.

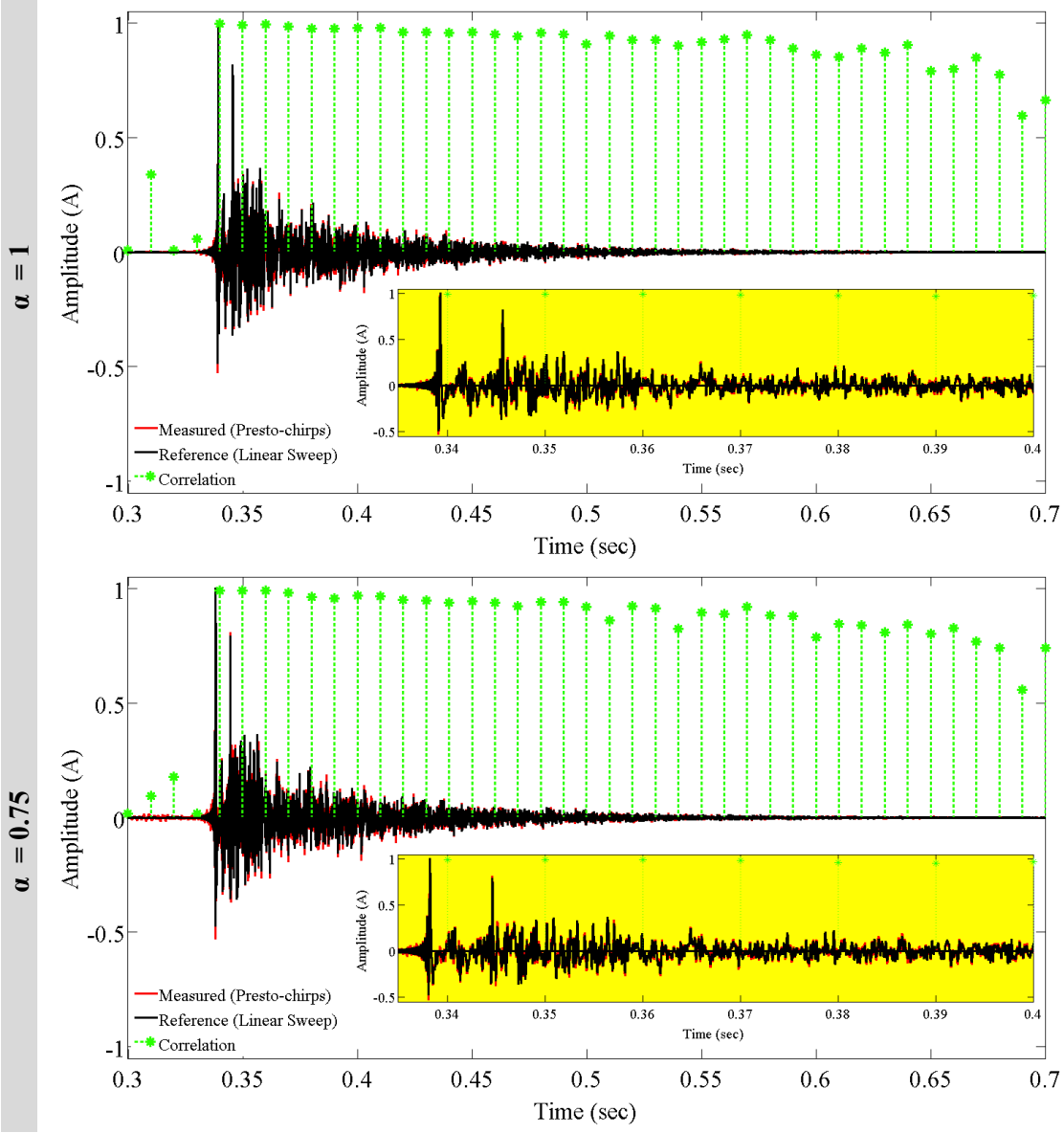
SNR [dB]

Bands [Hz]	31.5	63	125	250	500	1k	2k	4k*
LIN chirp	25.0	34.0	40.5	41.6	47.4	54.4	53.8	54.0
α = 1	39.5	40.3	42.8	45.6	45.2	51.5	50.5	50.7
α = 0.75	38.7	41.3	39.3	45.8	45.9	52.8	51.0	52.0
α = 0.5	32.4	34.3	43.8	43.5	46.6	50.9	49.6	52.8
α = 0.25	25.4	27.3	33.2	45.9	46.5	51.5	47.7	51.5
α = 0	28.4	28.0	25.4	29.5	26.5	32.7	38.8	39.0

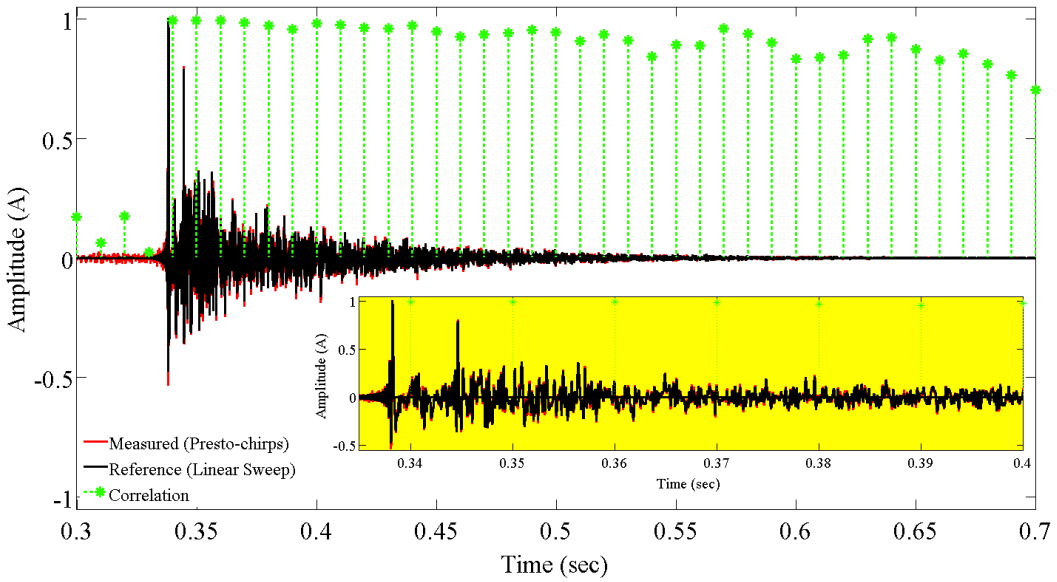
Table 19 – SNR given by the software Audacity. The first row, in blue, refers to the reference RIR measured using a 10 seconds long chirp (27 – 4.7 kHz).

8.3.1.1 Time Domain Analysis

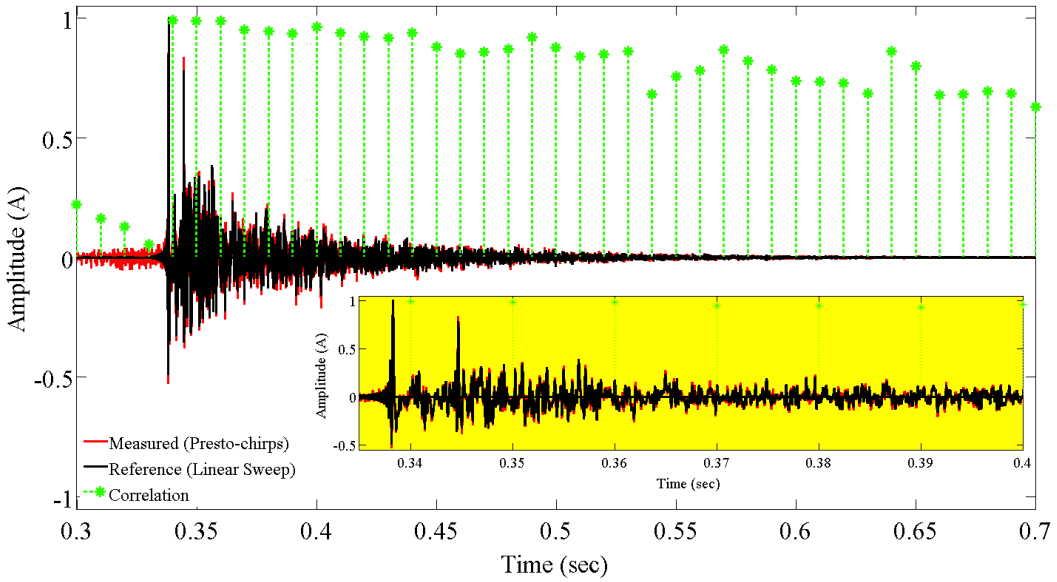
The time domain analysis (Figure 120) shows a good degree of correlation, mainly in the early part of the RIR, for the α=1 configuration. Starting from the tail, then spreading towards the main peak of the RIR, the correlation values (green marks) show a progressive loss of correlation as α decreases. This is due to the progressive increase of the synthesis noise along the RIR. The noise is clearly visible before the main peak, where it can also be seen that it is minimal for the α = 1.



α = 0.5


Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

α = 0.25


Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

α = 0

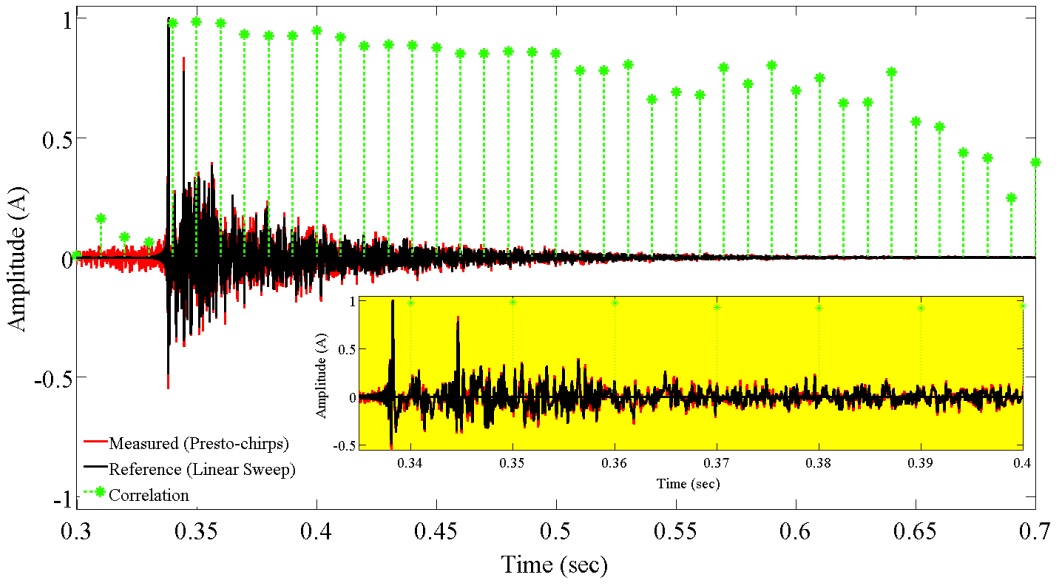

Amplitude (A)
Time (sec)
Measured (Presto-chirps)
Reference (Linear Sweep)
Correlation

Figure 120 - Comparisons between the reference RIR (in black) and the measured RIRs (in red). The green marks point out the degree of correlation between 10 ms long blocks of the two RIRs. In the yellow boxes are shown the first 40 ms of the RIRs.

Figure 121 shows the correlation of the whole reference RIR with those measured for the various α configurations. It confirms that correlation is best for α=1 and then progressively decrease as α decreases.

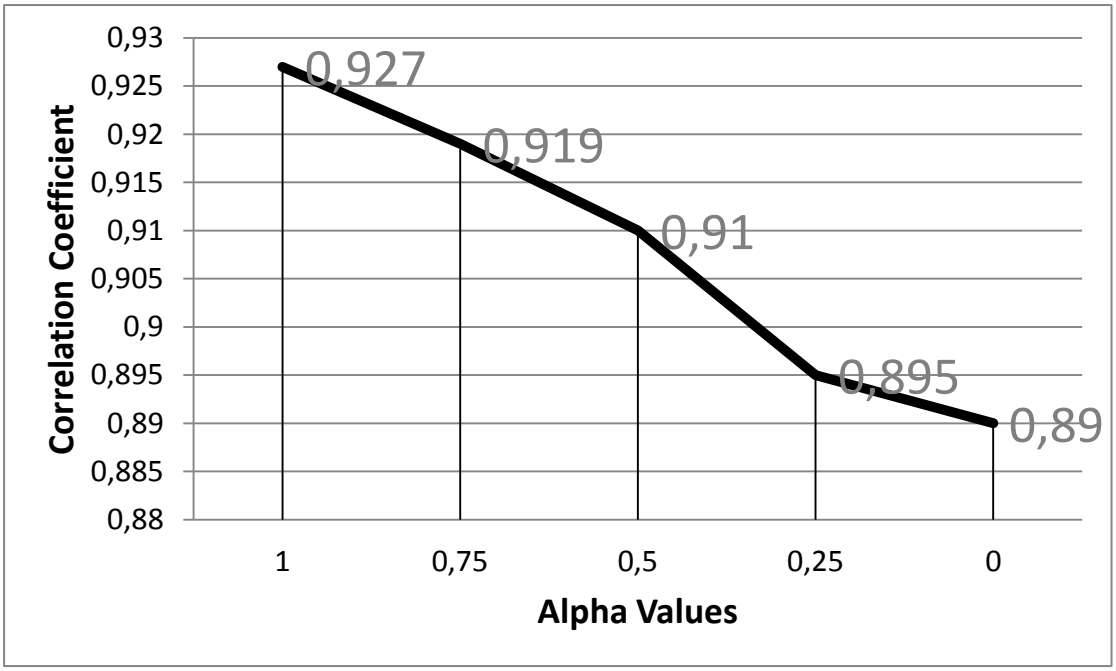
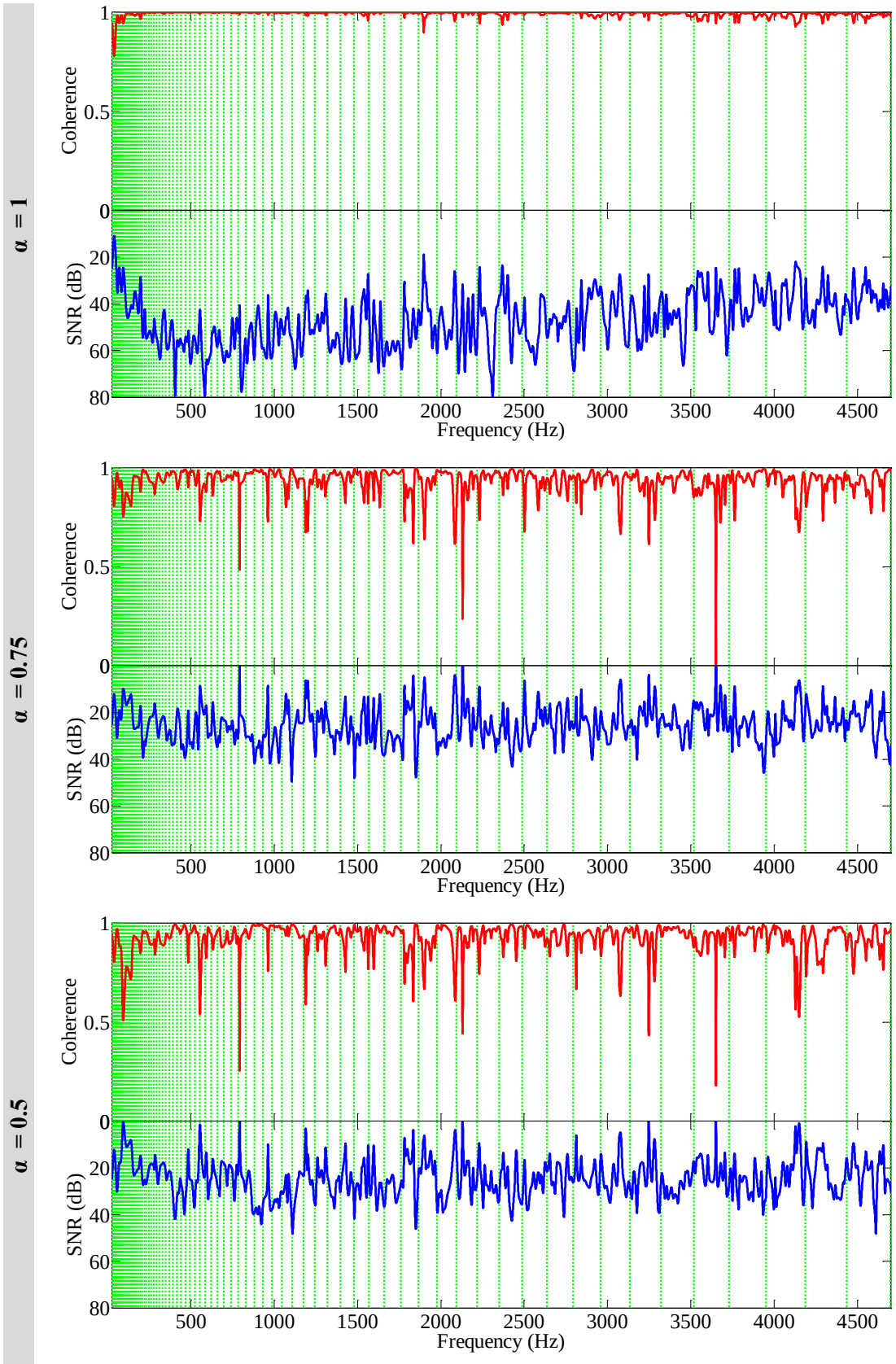


Figure 121 – Degree of correlation between the reference RIR and the RIRs measured with the different α-windowed stimuli.

Frequency Domain Analysis

Figure 118 shows the coherence and SNR functions, which both present the highest values for α=1. For lower α values there is a progressive diminution of both the coherence and SNR. The frequencies that show the lowest coherence, among the others, correspond to the frequencies having lower magnitude in the frequency response function of the room.


α = 1
α = 0.75
α = 0.5
Coherence
SNR (dB)
Frequency (Hz)
1
0.5
0
20
40
60
80
500
1000
1500
2000
2500
3000
3500
4000
4500

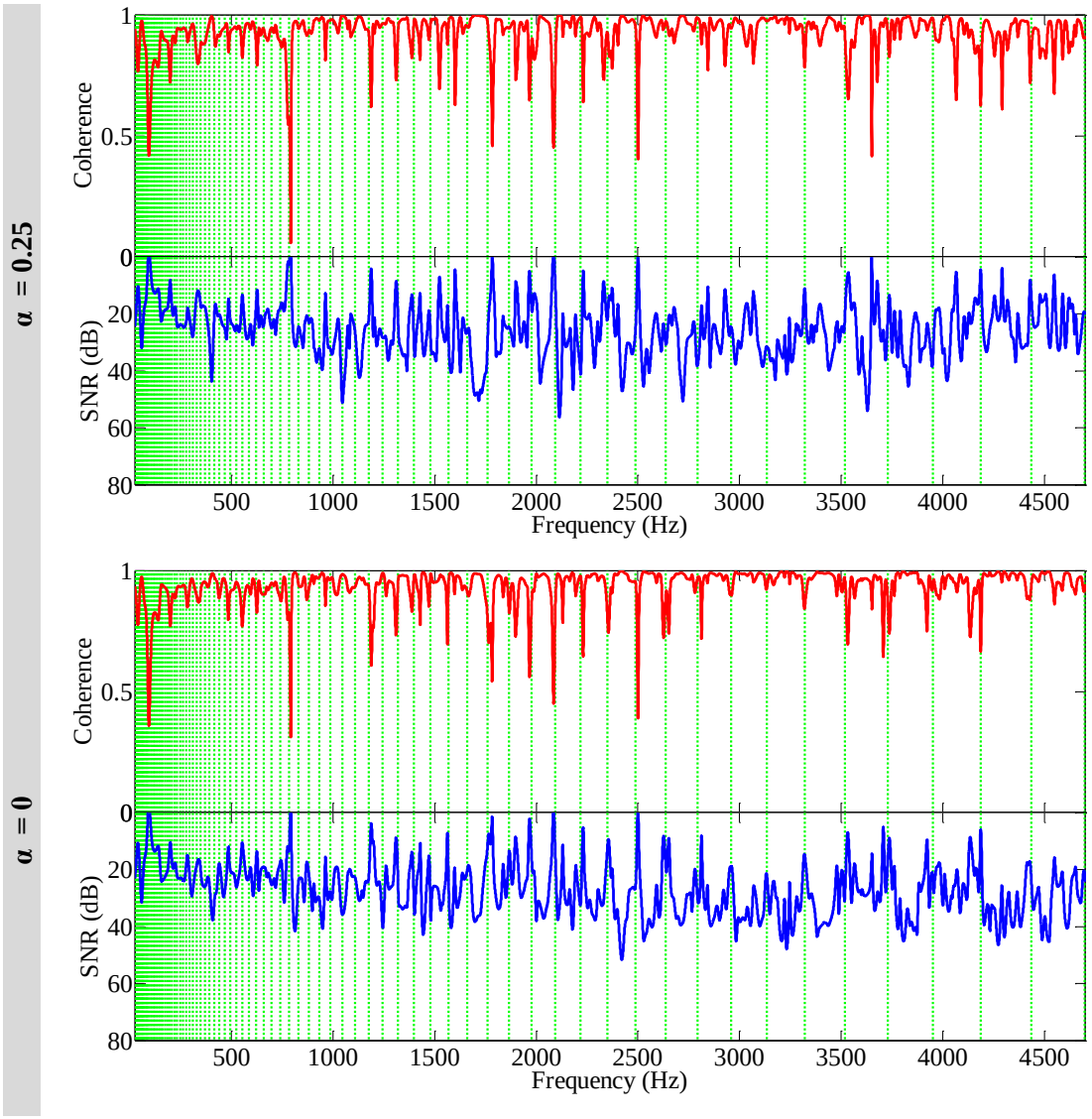


Figure 122 - In red is shown the coherence function, and in blue the SNR, between the reference FRF and the FRF calculated using different α related stimuli, as labelled. The dashed lines indicate the frequencies of notes.

Discussions of the Case Study

The presented case has given a useful insight into the accuracy of the method in dealing with short reverberation times. The accuracy is highest, both in frequency and in time domain, for α=1. The presence of the syntheses noise is consistent for α=0, which consequently alters the EDCs, and in turn the acoustic parameters. Similarly,

the coherence and the SNR functions show a progressively worse result as α decreases. Local minima of the reference FRF are first affected.

From the obtained results it can be stated that, for such a type of room, the presto-chirps method resolved to obtain a measure of the RIR with sufficient accuracy only for the α=1 configuration.