POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione

Master of Science in
Music and Acoustic Engineering

Perception of Initial Time Delay Gap, Reverberation time, and density of impulse response of a room

Candidate

Edoardo Epifani
Student Id. number 920668

Formal Thesis Supervisor

Prof. Augusto Sarti

Thesis Supervisor

Prof. Jędrzej Kociński

Academic Year 2020/2021

POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione

Laurea Magistrale in
Music and Acoustic Engineering

Percezione dell'Initial Time Delay Gap, Tempo di Riverberazione e densità di una risposta all'impulso di una stanza

Candidato

Edoardo Epifani
Matricola 920668

Relatore Formale

Prof. Augusto Sarti

Relatore

Prof. Jędrzej Kociński

Anno Accademico 2020/2021

Perception of Initial Time Delay Gap, Reverberation time, and density of impulse response of a room
Master thesis. Politecnico di Milano

This thesis has been typeset by LaTeX and the smcthesis class.

Author's email: edoardo.epifani@mail.polimi.it

Sommario

Questo studio di psicoacustica esplora come vengono percepiti tre parametri di un riverbero, l'Initial Time Delay Gap (ITDG), Reverberation Time (RT) e la densità delle prime riflessioni usando due tipi di campioni sonori e riverberi con lunghezze differenti per i primi due parametri citati.

Test creati appositamente su di un sito web con metodi adattativi usati per selezionare il tipo di suono adatto e metodi alternative forced-choice per fare in modo che il tester fosse nella migliore condizione per rispondere in base al tipo di test proposto.

L'analisi proposta mostra una forte correlazione tra la percezione dell'RT e la durata di un riverbero, cosa che non si può affermare per l'ITDG perché poco correlato. Per quanto riguarda il tipo di sample utilizzato, l'ITDG è lievemente influenzato rilevando così che un campione vocale ha ottenuto in media risultati minori, invece l'esito del test sulla densità dimostra che viene distinto più facilmente un riverbero con densità delle prime riflessioni più alta usando un campione strumentale, i risultati dei test dell'RT non rilevano alcuna differenza se non trascurabile.

Infine si può sostenere che i risultati ottenuti sono simili a quelli ottenuti in altri altri studi con un margine di errore maggiore causato dal fatto di non esser stati svolti tutti nelle condizioni ottimali e da persone esperte di ascolto musicale.

Abstract

This psychoacoustic study investigates how three parameters of a reverb are perceived, namely the Initial Time Delay Gap (ITDG), Reverberation Time (RT) and the density of early reflections using two types of sound samples and reverbs with different lengths for the first two parameters.

Tests were set up on a website with adaptive methods used to select the appropriate sound type and alternative forced-choice methods to ensure that the tester was in the best condition to respond according to the type of test proposed.

The proposed analysis shows a strong correlation between the perception of RT and the duration of a reverberation, which cannot be stated for the ITDG because it is poorly correlated. Concerning the type of sample employed, the ITDG is slightly affected, thus detecting that a vocal sample obtained on average lower results, while the outcome of the density test shows that a reverberation with a higher density of the first reflections is more easily distinguished using an instrumental sample, the results of the RT tests do not detect any difference if not negligible.

Finally, it can be said that the results obtained are similar to those obtained in other studies with a greater margin of error caused by the fact that they were not all carried out under optimal conditions and by people who are experts in musical listening.

Acknowledgments

Throughout the writing of this thesis I have received a great deal of support and assistance.

I would first like to thank my supervisor at Adam Mickiewicz University of Poznan, Professor Jędrzej Kociński, whose expertise was invaluable in formulating the research questions and methodology moreover your valuable guidance throughout my studies.

I would like to thank my formal supervisor, Professor Augusto Sarti, your insightful feedback pushed me to refine my thinking and took my work to the next level furthermore, I thank you for your immeasurable patience.

I would also like to thank my supervisor's colleague, Dr Jan Felcyn, who provided me with the tools I needed to choose the right direction with statistical analysis.

I would like to thank my colleagues at my university for their collaboration during the preliminary tests. Your advice was extremely valuable.

In addition, I would like to thank my friends, Dr. Natalia Giochev, Mattia Masiello and Chiara Zappa, for their precious guidance during the process of editing the different languages on the website and another thanks to my friend Ylenia Lauro for providing me with the most effective suggestions during the review of the thesis.

Finally, I could not have completed this thesis without the encouragement of my parents with their constant support. You are always there for me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There exist many different objective parameters describing the acoustical properties of a room. Nevertheless, one must have in mind that the final assessment is totally subjective. It is a listener who comes to a concert hall and with his/her hearing perceives the signals modified by a room with his/her hearing. That is why scientists still try to find the relation between objective parameters and subjective assessment.

This thesis analyses some of the relations in this matter. To receive this kind of results three psychoacoustics tests were prepared, one for recognizing the Just Noticeable Difference (JND) of Initial Time Delay Gap (ITDG), one for recognizing the JND in Reverberation Time (RT), and one for recognizing the JND in the density of early reflections of a reverberant room. Just noticeable difference JND is defined as the smallest perceivable change in a given measurable parameter and is an important factor that correlates the subjective dimension of sound perception to its objective measures.

Y. Ando was the father of all that we know about subjective judgment of acoustic parameters, but especially about concert halls as explained in his book [1]. Another who added much to this subject was L. Beranek, filling in some of the aspects not taken into account by Ando. There are similar studies in the literature for ITDG and RT, but they treat it differently using different evaluation methods, stimuli, and reverberations. On the other hand, the evaluation of the perception of the density of a reverberation does not include many studies on the subject, and these are not related to the case in this thesis, although some methods of comparison do exist and will be taken into account.

For the study of ITDG and RT, we use as reference 4 different RTs of 0,5 s, 1 s, 2 s, and 4 s to understand better the differences in perception for different durations combining adaptive methods with forced-choice alternatives to achieve optimised values. Moreover, since the test was carried out via the Internet, we ask every tester to reply to simple questions about the condition in which they did the tests so that it is possible to make observations on those as well. Using this study, understanding what are the JNDs for different properties of room for these three objective parameters of the enclosure acoustic field can be useful to improve the possible models of sound perception in reverberant conditions. Improve the design accuracy of concert halls, ray tracing, and sound localization.

Therefore the thesis is arranged as follows: first, the state-of-the-art in which all the theory needed to understand the three types of parameters tested will be described and the studies already carried out on them. This is succeeded by the chapter on goals in which it is explained what the aims are and anticipates the methods used to achieve them which are in the next chapter, in which there are insights to understand how the tests were done and why the theoretical and practical choices were made, starting from the methods to the creation and selection of reverbs up to the details of the website. Then comes the chapter on results with tables and graphs to make them easier to read and understand, followed by the discussion part, in which the results are analysed and compared with the literature. Conclusions and future work is the last chapter of this thesis in which it summarises what was done in the thesis and why what were the results and the most important conclusions and finally what could be done in the future on this topic.

Chapter 2

State of the Art

2.1 Geometrical theory in room acoustics

In the Greek and Roman open-air theatres of antiquity, the sound level increase by a single reflection of the ground had a very favorable effect. However, in enclosed spaces, such as concert halls, sound interacts in different ways with the various surfaces of the surrounding architectural environment, imposing different amplitudes, different directions of propagation, etc., on the reflections. The result of this interaction, from the point of view of a listener in the audience or a musician on stage, is a series of successive sound waves arriving after the direct sound with different amplitudes, time delays, and directions. This time response is extremely important in assessing the acoustic quality of a room, as it represents the room's acoustic "signature" that is call the room impulse response (RIR). Note that there is an impulse response (IR) for each position of the source and receiver under consideration.

The first wavefront that reaches a listener without interaction with any obstacles or boundaries is called direct sound. In an enclosed space, the sound wave propagating in all directions will sooner or later interact with a surface and, consequently, be reflected Figure 2.1.

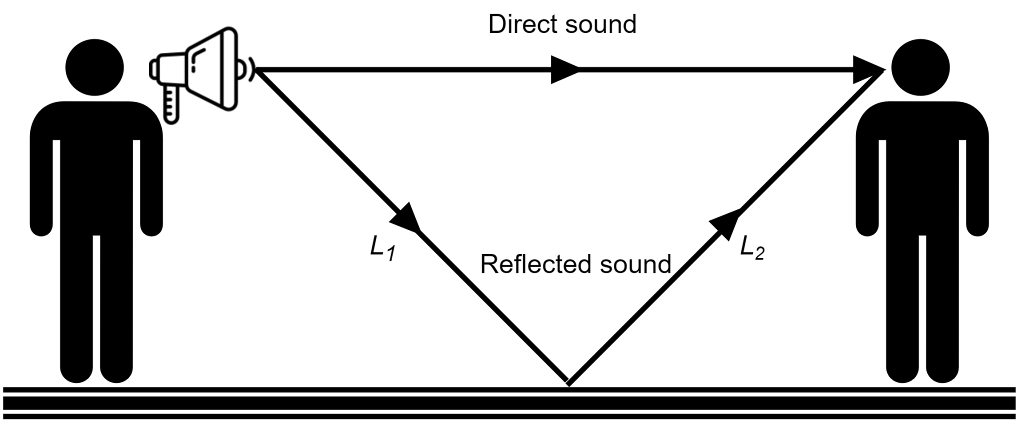


Figure 2.1. Example of a direct sound and only one reflected sound.

Consider the case illustrated in Figure 2.1. The reflection from the ground reaches the listener after the direct sound with a time delay of equal:

$$\Delta t = t_1 - t_0 = \frac{(L_1 + L_2) - L}{c} \tag{2.1}$$

where t_0 and t_1 are the times of arrival of the direct and reflected sound waves, respectively, and c is the speed of sound (~ 344 m/s). Although simple, the Equation (2.1) is extremely important in predicting how this sound reflection affects the listener's perception of the listener. If the reflected sound arrives within a short time interval after the direct sound (≤ 50 ms, explained in Chapter 2.4), our ear will integrate both sounds as one and add the corresponding intensities. However, larger time delays will cause us to understand two separate sounds (echo).

Figure 2.2 schematically represents the arrival times of direct and reflected sound waves with their corresponding amplitudes. This representation, also called an echogram, shows that the reflections (R_i) have lower amplitudes than the direct sound (D) due to the longer propagation path and the interaction with the room surfaces.

The Initial Time Delay Gap (ITDG) is the time delay between the direct sound and the first reflected sound wave arriving at the receiver and is extremely important in defining the acoustic quality of the room.

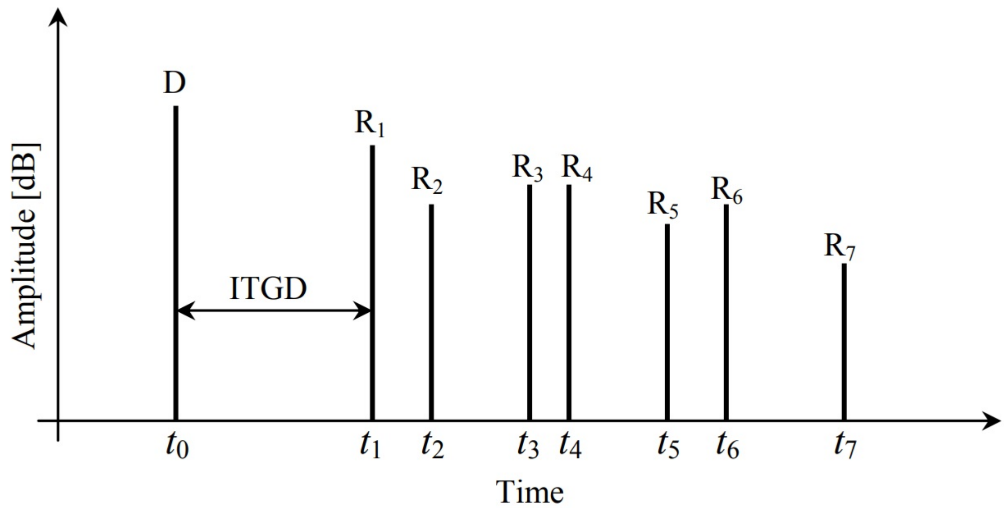


Figure 2.2. Echogram with early reflections and the Initial Time Delay Gap (ITDG) highlighted.[2]

The previous Figure 2.2 shows the early sound reflections, which in large halls usually arrive within 80-100 ms of the direct sound, mainly due to the first reflections from the ceiling, floor, or walls. However, the reflection pattern is much more complicated and has such a high reflection density, especially after the early sound, that individual reflections can no longer be distinguished. These late reflections from the so-called reverberation field. Its main characteristic is that with simple room geometry and diffuse field conditions, the corresponding sound pressure drops exponentially, with the reverberation time defined as the time it takes for these reflections to become almost inaudible (corresponding to a sound level decrease of 60 dB explain in Chapter 2.3).

When more accurate calculations are required, a geometric approach is usually used. This approach describes sound propagation as sound rays, which can be a valid assumption if only wavelengths smaller than the characteristic dimensions of space, surfaces, and obstacles are considered. Although this is a technique mainly used for numerical calculations, a revised theory has been developed to describe the decay of sound in rooms, where it was found that the reverberation field is not constant, but steadily decreases with increasing distance from the sound source. [3]

In real rooms, the reverberant field is only approximately diffuse, with reflections from multiple directions arriving at most listener positions with a similar pattern. However, the degree of diffusivity depends on the uniformity of the distribution of absorption surfaces in the room. Thus, as seen, the received sound in a room can be divided into three components: Direct sound, early reflections, and late reverberant sound.

After the early reflections, the exponential decrease of sound pressure over time turns into a linear decay when considering the corresponding logarithmic quantity, the sound pressure level. This (approximately) linear decay. The presence of the early reflections, which gradually influence the decay process, is visible between 50-80 ms. [2]

2.2 Echogram

The simplest way to idealize a reverb is with the echogram that is a temporal distribution of reflections caused by the bouncing back of the sound from the walls, ceiling, and ground in a room.

If the absorption coefficients of all walls are frequency independent, the received signal $s'(t)$ is the superposition of infinitely many replicas of the original signal [4], each of them with its particular strength A_n and delayed by its particular traveling time t_n:

$$s'(t) = \sum_n A_n s\left(t - t_n\right) \tag{2.2}$$

Accordingly, the impulse response of the room reads in Figure 2.3:

$$g(t) = \sum_n A_n \delta\left(t - t_n\right) \tag{2.3}$$

Marking the arrival times of the various reflections by vertical lines on a horizontal time axis and choosing the height of the lines proportional to the relative strengths of the reflections, i.e., to the coefficients An, is obtained what is often called a *"reflection diagram"* or *"echogram"*. It contains all the essential information about the temporal structure of the sound field at a given point in space. In Figure 2.3 a schematic reflection diagram is plotted. After the direct sound arriving at $t = 0$, the first strong reflections occur sporadically at first, but later their temporal density increases rapidly; at the same time, the reflections carry less and less energy. The role of the first isolated reflections concerning our subjective hearing impression is quite different from that of the very numerous weak reflections arriving at later times, which merge into what is subjectively perceive as reverberation. Thus it is possible to consider the reverberation of a room not only as of the common effect of freely decaying vibrational modes but also as the sum of all reflections - except the very first ones.

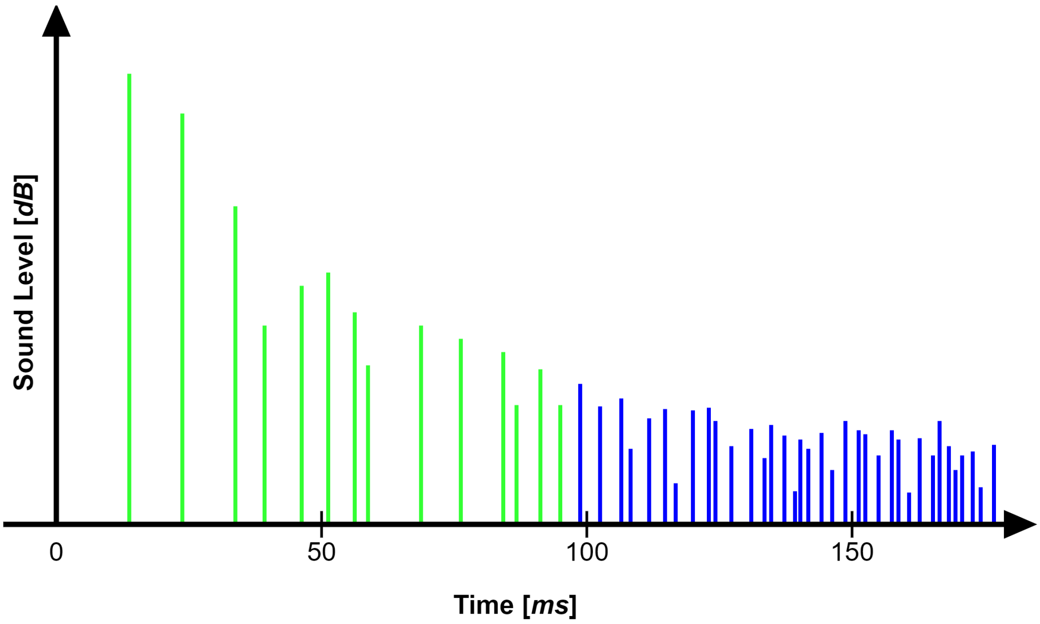


Figure 2.3. Example of echogram with early reflections (green lines) and late reflections (blue lines).

2.3 Reverberation Time

Reverberation time is one of the main acoustic requirements which contribute to the well-being of the occupants of an environment, especially in those work environments where listening and verbal communication are required.

In a reverberant field, if a sound source suddenly stops to emit, the sound still does not stop, but persists for a certain time, thanks to the reflections on the surfaces (which are called reflected echoes).

The decay time, also called *"Reverberation Time"* (RT), depends on the speed of sound, the volume of the enclosure, the distance between the walls, the number, the quality, and the sound absorption capacity of the reflecting surfaces.

The term *"Reverberation Time"* is defined as the necessary time to obtain a decay of 60 dB of the sound level, starting from the moment of interruption of the sound source.

In environments with highly reflective walls, such as classrooms or canteens not treated with sound-absorbing materials, the reverberation time is long, whereas, in environments with walls which are covered with highly sound-absorbing materials, the reverberation time is reduced. The reverberation time must be adequate for the intended use of the environment. A very long reverberation time causes loss of speech intelligibility and increases background noise.

2.4 Haas effect (the precedence effect)

The Haas effect is a psychoacoustic effect described for the first time by Helmut Haas in his 1949 dissertation [5], but actually discovered by Lothar Cremer the previous year and called “the law of the first wavefront” published in 1977 [6]. It is also known as the underlying precedence effect.

The precedence effect is a binaural psychoacoustic effect, which means that when a sound is followed by another sound divided by a proper short time delay (under the listener’s echo threshold), the listener perceives a single auditory event. The perceived spatial position is dominated by the position of the first-arriving sound (the first wavefront). The lagging sound also affects the perceived position, nevertheless, its effect is inhibited by the first-arriving sound.

The precedence effect occurs when the lagging wavefronts arrive between 2 ms and about 50 ms later than the first wavefront. This range is signal-dependent. In speech, the precedence effect disappears at delays longer than 50 ms, while for music, the precedence effect can occur at delays as short as a few 100 ms.

In two-click delay experiments, localization effects include aspects of summing localization, localization dominance and suppression of delay discrimination.

- Summing localization: for time delays less than 2 ms, listeners perceive only one tone; its direction lies between the positions of the lead and lag tones. One use for summing localization is intensity stereophony, in which two loudspeakers radiate the same signal at different levels, resulting in a localized sound direction among the two loudspeakers. The localized direction hangs on the level difference between the loudspeakers.

- Localization dominance: for delays between 2 and 5 ms, listeners also perceive a sound whose location is determined by the location of the leading sound.

- Lag discrimination suppression: at short time delays, listeners are less able to discriminate the location of the lagging sound.

The last two are generally considered aspects of the precedence effect.

Regarding Time delays longer than 50 ms (for speech) or just over 100 ms (for music), the delayed sound is perceived as an echo of the first-arriving sound. Both sound directions are correctly localized. The time delay for distinguishing echoes differs on the signal characteristics. Thus, as regards signals with impulse characteristics, echoes are only perceived after a delay of 50 ms, while for signals with almost constant amplitude, the echo threshold can be increased up to time differences of 1 to 2 s.

A special manifestation of the precedence effect is the Haas effect. Haas showed that the precedence effect also occurs when the level of the delayed sound is up to 10 dB higher than the level of the first wavefront. In this case, the range of delays in which the precedence effect operates is reduced to delays between 10 and 30 ms.

2.5 Relation between subjective and objective parameters

In the past, researchers did tests on several parameters of reverbs to investigate the factors most important for determining the acoustic quality of rooms, in particular, Concert Halls with the objective and subjective preference of audience in Concert Halls. Beranek and Ando were the two who gave the greatest contribution to this study.

Ando's approach differs from other work on subjective preference in one important aspect. He starts by characterizing the source in terms of the autocorrelation function (ACF) of the music signal. [7]

In Ando's book [1] audience preference is then shown to depend upon four objective and independent parameters, three of which are "monaural-temporal" quantities while the fourth is "binaural-spatial" and with all these four parameters determine the scale values of subjective preference for simulated sound fields. The three monaural-temporal quantities are the listening level (LL), the initial time delay gap (ITDG), the subsequent reverberation time (T_{sub}), and the criteria for these quantities are all expressed in terms of the source ACF. The binaural-spatial quantity is expressed in terms of the interaural cross-correlation coefficient (IACC). No attempt is made to relate these quantities to the usual subjective factors such as "clarity," "warmth," "envelopment," and so on, as the origin of the objective parameters in unqualified paired comparison preference experiments avoids the need for these troublesome terms.

Beranek suggested adding two more factors to Ando's four physical factors, the Bass ratio (BR) and the Surface Diffusivity Index (SDI) [8]. Problems with regard to both objective and subjective orthogonalities in the above two investigations remain. As far as subjective preference is concerned, it is worth noting that the sound fields in a real concert hall for listeners facing the performers (IACC is obtained at $T = 0$) can be described by only four orthogonal physical factors. [9] [10]

2.5.1 Listening Level (LL)

The listening level is, of course, the primary criterion for listening to the sound field in concert halls. The absolute preferred listening level depends upon the musical program and the particular passage being performed. The LL is given by

$$\mathrm{LL} = 10\log\left(1 + \mathrm{A}^2\right) - 20\log \mathrm{d}_0 - 11[\ \mathrm{dB}] \tag{2.4}$$

where A is the total pressure amplitude of the early reflections and subsequent reverberation, and $d_0(= |r - r_o|)$ is the distance between the source and the listener's position. [11]

2.5.2 Initial Time Delay Gap (ITDG)

The relationship between the autocorrelation function (ACF) of the source signals, the total pressure amplitude of the early reflections and subsequent reverberation, and the most preferred delay time of a single reflection has been obtained from the approximate formula [1] [12]

$$[\Delta t_1]_p \approx (1 - \log_{10} A)\, \tau_e \tag{2.5}$$

where the total amplitude of the reflections A is given by

$$A = \sqrt{A_1^2 + A_2^2 + \cdots + A_N^2} \tag{2.6}$$

$A_i (i = 1, 2, ..., N)$ being the pressure amplitude of each reflection relative to that of the direct sound, N is selected a large number for the convergence. The quantity τ_e is the effective duration of the ACF with a 10-percentile delay, defined as the delay at which the envelope of the normalized autocorrelation function becomes 0.1.

2.5.3 Subsequent reverberation time

For flat frequency characteristics of reverberation, the preferred subsequent reverberation time after early reflections is described simply in terms of the effective duration of ACF of the source signals [13] [14] , as given by

$$[T_{sub}]_p \approx 23\tau_e \tag{2.7}$$

2.5.4 Interaural Cross-Correlation Coefficient (IACC)

The IACC and subjective preference show a negative correlation for all available data. That is, listeners prefer dissimilar signals for the left and right ears. This holds only under the condition that the maximum value of the interaural cross-correlation function at an interaural time delay equal to zero is maintained to ensure frontal localization of the source [10]. Otherwise, an image shift of the source or an unbalanced sound field may be perceived, and the value of preference decreases.

2.5.5 Bass Ratio (BR)

In the large halls, adequate strength of the bass sounds is difficult to achieve because it depends on the materials used in the construction of the walls, ceiling, and floor, on the thickness of the upholstering on the chairs on carpets, or other sound-absorbing materials used for echo control, and on ventilation and lighting fixtures and other openings. [15]

L.L. Beranek defined a corresponding BR

$$BR = \frac{T_{125} + T_{250}}{T_{500} + T_{1000}} \tag{2.8}$$

with reverberation times at 125 and 250, respectively 500 and 1000 Hz (in an occupied hall) as an acoustical quality criterion [16]. Accordingly, this should assume values between 1,1 and 1,5 for rooms with a generally high, respectively low mean reverberation time Tm: "If the surfaces of the walls or ceilings or seats absorb the low frequencies, the full orchestra may sound deficient in basses and cellos... A hall lacks warmth when the reverberation times are lower at low frequencies (75 to 350 Hz) than at mid-frequencies (350 to 1400 Hz), i.e. low BR". [8]

When looking through the later literature, one may find statements that sound very logical, though only from an energetic point of view [17]:

- All human voices and musical instruments, including a full orchestra sound, radiate a long-time averaged sound spectrum that drops toward low frequencies. Therefore somewhat stronger reflections from the room boundaries should be beneficial for its low-frequency part and hence for the fullness and warmth of the sound.
- The human ear is less sensitive, respectively the hearing threshold is higher toward low frequencies, in fact, the more so, the less loud the performance is. For this reason, it should be expedient, particularly in large rooms, to support the sound through a stronger reverberation.
- Curves of equal loudness move closer and closer together toward low frequencies. For all frequencies to remain audible equally long during pauses, the sound level decay should be delayed by a longer reverberation of the low than of the high frequencies.

2.5.6 Surface Diffusivity Index (SDI)

Every successful concert hall appears to have an abundance of surface irregularities. These irregularities diffuse the acoustical energy in the room and lend a smooth, homogenized feeling to the early sound and the late decaying sound. These often take the form of coffers in the ceiling, and niches, statues, and baroque decorations on the sidewalls. The SDI is determined from visual inspections of photographs or visits to the concert halls. The highest-rated halls have SDI's of 1.0, while those of the lowest-rated halls fall in the range of 0.3 to 0.7. The intermediate halls have SDI's of the order of $0.7 + 0.1$. [15]

2.6 Reverberation Time assessment

Given the large number of acoustic parameters which can be measured nowadays and their widespread usage, it is particularly useful to have well known Just Noticeable Difference (JND) values for each one of them. [18]

The first studies carried out into this field were on Reverberation Time (RT), as reported by Cremer and Muller [19], showing that for RT values above 0.6 s the JND was about 4% of the observed value, while for those below 0.6 s the subjective limen became "absolute" and approximately equal to 0.024 s. Above 0.6 s, the relative variation JND was strictly dependent on RT values, showing a U-shaped trend with high values at short RT, reaching a minimum of about 4% between 1 s and 2 s, and then increasing again up to 5% when RT approached 6 s. More recent studies [20], based on realistic binaural reproduction of sound fields, have shown that below 0.6 s the difference limen was 0.042 ± 0.015 s, twice as the value resulting from previous studies. However, significant variations have appeared as a function of the program material used, suggesting that the transient nature of the signals may explain such variance. [18]

In the study of 2013 [21] they tried to quantify the JND of RT with three different length of RT (1, 2, and 3 s), a band-limited noise centered around 1000 Hz using the "one-up, two-down" method Chapter 4.1 and a three-alternative forced-choice paradigm employed. Testers had a minimum of 3 years of musical training or experience and the test was conducted in an acoustically-treated listening booth with neutral-frequency headphones where the sound level was calibrated at 75 dBA and constant for all the subjects. Researchers performed these tests with steps below and above the RT which was taken into account and stopped the test after the fifth reversal (turning point) of the adaptive procedure and found out that "The overall JND of RT has been found to be 24.5% using 1000 Hz octave-band limited noise for RTs between one and three seconds." [21] with a standard deviation of 6.09%.

According to Hak [22], the Just Noticeable Difference (JND) for the RT is approximately 5% and ISO 3382-1 [23] lists the JND of reverberation metrics at 5% based on work by Seraphim (1958) [24]. However, other researchers have found the JND of RT to be higher from 6% to 39%. Many of these studies utilized band-limited stimuli, such as speech, music motifs and bandlimited noise.

2.7 Initial Time Delay Gap assessment

In relation to the study of J. Atagi, R. Weber, and V. Mellert [25] made starting from the Ando's one, researchers have carried out tests to find out that the JND of a single reflection fixed at 240 ms is equal to 10.6 ms using as sound source a musical motif (Royal Pavane by Gibbons, 10 s). Slightly different results are expected from the tests done in this study because instead of using a RIR they used a single reflection that is placed at 240ms.

"Reichardt et al. (1967) showed that the JND of initial-time delay gap for the lateral reflection is 7 ms $\pm$ 0.6 ms." [26]. Of all studies, this study is the most similar one to what has been done in this thesis about the perception of the ITDG.

2.8 Echogram density assessment

As concerns literature, there are no assessments for this kind of tests because it is not a conventional parameter of reverb, but there are similar studies about that. Hawkes and Douglas [27] identified five parameters which are five dimensions that have been expanded and refined since that time. These parameters are clarity, reverberation, envelopment, intimacy, and loudness. Excluding the first reflection (ITDG) which describes the intimacy index, the density of the other early reflections can partially describe the clarity index which is defined as "The degree to which rapidly occurring individual sounds are distinguishable." [28] described by three Acoustical Factors: RT, Early-to-late energy ratio, and speed of the music.

The energy-to-late ratio is defined by

$$C_{t_e} = 10\log_{10}\left(\int_0^{t_e} p^2(t)dt \Big/ \int_{t_e}^{\infty} p^2(t)dt\right) [dB] \tag{2.9}$$

where $p^2(t)$ is the squared room impulse response and t_e is a characteristic critical delay time in milliseconds. C_te is a measure of the energy ratio between early and late reflections and therefore low for highly reverberant rooms and vice versa. [29]

A further essential factor which characterizes the intelligibility of the sound, especially as concerns the speech, is the parameter "definition" D:

$$D = \frac{\int_0^{50\ \mathrm{ms}} p^2(t)\mathrm{d}t}{\int_0^{\infty} p^2(t)\mathrm{d}t} \tag{2.10}$$

similar to clarity C but having both integrals starting at 0 and the integral at numerator being a subset of the second, the result will always be less than or equal to 1.

In the ISO [30] the clarity index with $t_e = 80ms(C_{80})$ has the JND of 1 dB and the JND for D is 0.05.

Chapter 3

Goals

The goals of this thesis is to discover how much humans are able to distinguish the perception of three properties of reverbs, Initial Time Delay Gap (ITDG), RT, and density of early reflections.

By deciding to use different RTs for the ITDG and RT tests, one expects to obtain differences in the final results, particularly for the latter RT tests mentioned in the Chapter 2.6 whose results are that the Just Noticeable Difference (JND) will be more or less the 25% of the RT. In the tests made in this thesis the procedures are different and also the stimuli thus obtaining different results.

The ITDG test is expected to be a slightly different perception changing the RT because the differences between the reverb of 0.5 s and the reverb of 4 s are huge and that is the reason why the tester will have some difficulties recognizing them.

To obtain the results that have been used in this thesis, people of any age and gender were asked to do the test, therefore it will not be possible to conduct a precise study based on these characteristics, because for each test there will not be balanced results, however some consideration will be made.

An adaptive 1up-1down procedure was used to obtain the best results from such tests because as explained in the Chapter 4.1 from the point of view of the trade-off between the actual estimation of the result and the total testing time it is the most suitable one compared to a 1up-2down or 1up-3down, even if they are more precise.

As for the method of choice, the testers used in these tests were the 2AFC and 3AFC explained in the Chapter 4.2, which are simple methods, but very effective. Concerning the density test, which is the most difficult to understand, for testers who do not have experience listening to music, we decided to use the 3AFC so that even if we did not understand perfectly which characteristic to listen to, among the three sounds proposed, we will choose the one that is perceived by the ear as less similar. With regard to the other two tests, ITDG and RT, we used 2AFC because, unlike the other test, these two have, from a physical and practical point of view, very perceptible acoustic differences.

For a better and deeper understanding of the above-mentioned problems, some basic theories and related experiments must be described.

Chapter 4

Methodologies

4.1 Adaptive up-down procedures

4.1.1 The Simple Up-Down or Staircase Method

A relatively efficient method of estimating the 50% level is the simple up-down or staircase method [31]. It is similar to the method of limits in that the stimulus level is decreased after a positive response (or increased after a negative response), but unlike the method of limits, the test is not terminated after the first reversal. It is recommended that the test continues until at least six or eight reversals are obtained (Wetherill and Levitt, 1965). A typical data set is shown in Figure 4.1.

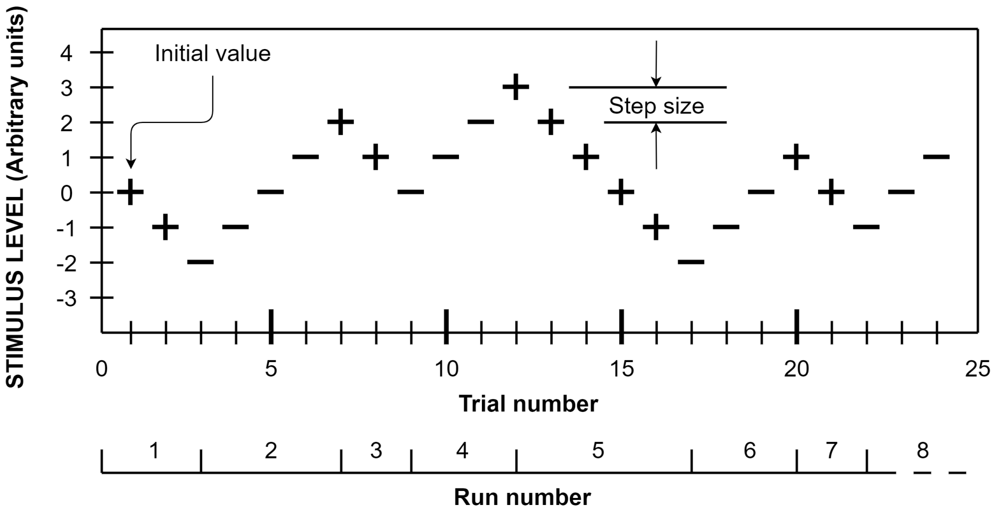


Figure 4.1. Example of 1-up 1-down procedures with constant step. The initial value is typically the best a priori estimate of X_{50}.

The increments by which the stimulus is either increased or decreased are called steps. For the experiment shown in Figure 4.1, a constant step size was used throughout. A series of steps in only one direction is defined as one run. Thus, in the example given, trials 1 through 3 constitute the first run, 3 through 7 constitute the second run, 7 through 9 constitute the third run, and so on. The level of the very first trial is the initial value.

The simple up-down procedure has the following advantages. Most observations are placed at or near X_{50}, which is a good placement of observations for estimating the 50% level. Second, if there is a gradual drift during the test, the placement of observations will follow this drift. However, the technique has the following drawbacks. The data are not well placed for estimating points other than X_{50}. Second, difficulties occur with very large or very small step sizes. If too large a step size is used, some of the data will be poorly placed relative to X_{50}. If too small a step size is used, then many observations will be wasted converging to X_{50}. This can be a serious problem if a poor initial value is used. A third shortcoming is a feature of psychophysical tests, since the subject, knowing that a sequential rule is being used, can anticipate the stimuli and adjust their responses accordingly.

Entry	UP group increase level after:	DOWN group decrease level after:	Probability of a sequence from DOWN *group* = $P[DOWN]$	Probability of positive response at convergence
1	$-$	$+$	$P(X)$	$P(X) = 0.5$
2	$+-$ or $-$	$++$	$[P(X)]^2$	$P(X) = 0.707$
3	$--$	$-+$ or $+$	$[1 - P(X)]P(X) + P(X)$	$P(X) = 0.293$
4	$++-$ or $+-$ or $-$	$+++$	$[P(X)]^3$	$P(X) = 0.794$

Table 4.1. Response groupings for transformed up-down strategies. Several simple response groupings are shown. Entry 1 corresponds to the simple up-down procedure. Entry 2 corresponds to the method used by Zwislocki (1968) and Heinemann (1961). Entries 2 and 3, with random interleaving, were used by Levitt (1964). [31]

4.1.2 The Transformed Up-Down Procedure

The simple up-down procedure is designed primarily to place observations in the region of X_{50} and is therefore not well suited to estimating points outside X_{50}. A general method for estimating points on a psychometric function is the transformed up-down procedure. The observations or sequences of observations are divided into two mutually exclusive groups. These are referred to as group UP and group DOWN respectively. The method used to group the observations depends on the point to be estimated. Table 4.1 shows some typical groupings, including some that have been proposed in other contexts (Zwislocki et al., 1958; Heinemann, 1961; Campbell, 1963). The stimulus level control rule is analogous to the simple up-down rule, except that the stimulus level is not changed until there is a sequence of observations belonging to either the UP or the DOWN group. The stimulus level is not changed until such a sequence is obtained. For example, according to Entry 2 in Table 4.1, the stimulus level would be increased either after a negative response or after a positive response followed by a negative response on the next trial. The stimulus level is decreased after two consecutive trials yielding positive responses. Note that as the test progresses, one sequence or the other must be achieved. Some typical data for this strategy are shown in Figure 4.2.

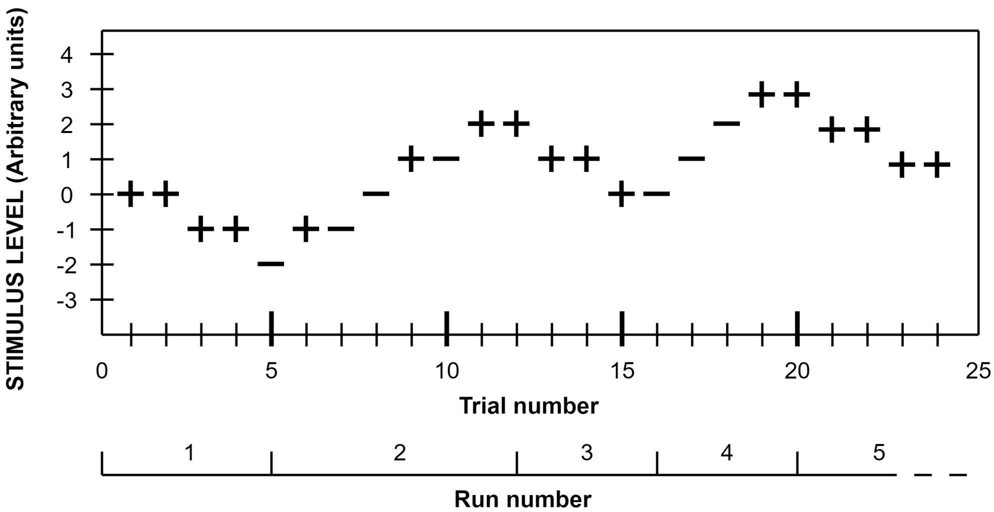


Figure 4.2. Example of 1-up 2-down procedures with constant step. The data are representative of the theory for achieving the $X_{70.7}$ level, described in Entry 2 of Table 4.1.

In this illustrative example, the changes in stimulus level are similar to those for the simple up-down strategy shown in Figure 4.1. That is, in Figure 4.2, if the - and + - response sequences belonging to the UP group are replaced by a single - response and the + + sequence belonging to the DOWN group is replaced by a single + response, then the resulting set of transformed responses is identical to that for the simple up-down strategy in Figure 4.1. However, in the case of the transformed up-down strategy, the average number of trials per run is increased. For example, runs 2, 3, and 4 consists of trials 5-12, 11-16, and 15-20, respectively, and the mean values of these runs are 0, 1, and 1.5 units, respectively. The transformed up-down strategy tends to converge on the stimulus level where the probability of a DOWN response sequence equals the probability of a UP response sequence (i.e., the probability of both equals 0.5). It is relatively straightforward to calculate the probability of obtaining either a UP or a DOWN sequence. The probability of obtaining a UP sequence (i.e., either +- or -) is $P(X)[1-P(X)]+[1-P(X)]$, where $P(X)$ is the probability of a positive stimulus-level X response. The probability of obtaining a DOWN response sequence (i.e., ++) is simply $[P(X)]^2$. Thus, the strategy converges on the value of X where $[P(X)]^2 = 0.5$, i.e., $P(X) = 0.707$. [31]

4.2 Alternative forced-choice methods

4.2.1 Two-alternative forced-choice

Two-Alternative-Forced-Choice (2AFC) is a method of psychophysics developed by Gustav Theodor Fechner [32] and is used to measure the sensitivity of a person, child or infant, or animal to a particular sensory input, a stimulus, through that observer's pattern of choices and reaction times to two versions of the sensory input. For example, to determine a person's sensitivity to dim light, the observer is presented with a series of trials in which a dim light was randomly displayed at either the top or bottom of the display. After each trial, the observer replies "top" or "bottom" the observer is not authorized to say "I do not know" or "I am not convinced" or "I did not see anything". In this sense, the observer's choice between the two alternatives is forced.

Both options can be presented simultaneously (as in the example above) or sequentially in two intervals (also known as two-interval forced-choice, 2IFC). For example, to determine sensitivity to dim light in a two-interval forced-choice procedure, an observer could be presented with a series of trials consisting of two sub-trials (intervals) in which the dim light is randomly presented in the first or second interval. After each trial, the observer answers only "first" or "second" as in an example of the site created for this study in Figure 4.3.

The term 2AFC is often incorrectly used to describe a yes-no task in which the observer is presented with a series of trials in which a stimulus is randomly presented on some trials and not on others. The observer only responds "yes" or "no" after each trial. The conclusions of a yes-no task are much more likely to be influenced by various response predispositions than in 2AFC tasks. For example, given an extremely dim light, a person might answer "No" completely truthfully (i.e., "I did not see any light") on each trial, whereas the results of a 2AFC task show that the person can reliably determine the position (top or bottom) of the same extremely dim light.

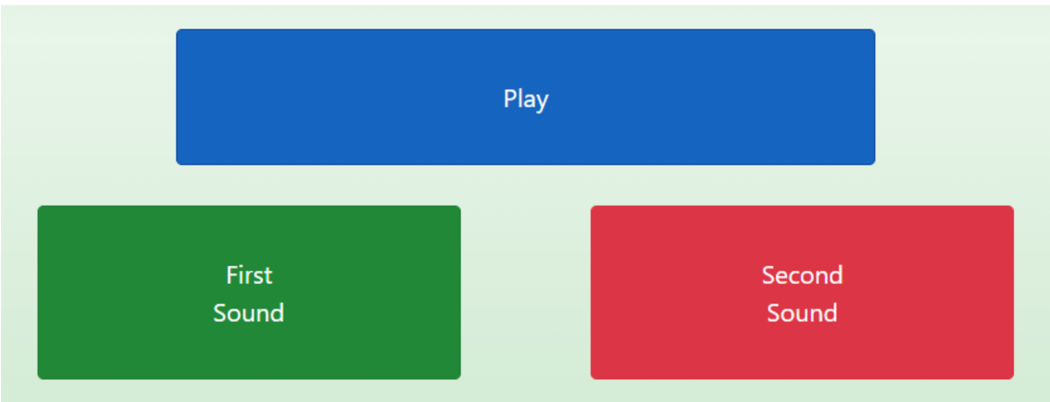


Figure 4.3. Example of 2AFC in the website after selecting the sound.

4.2.2 Three-alternative forced-choice

The three-alternative forced-choice (3AFC) is a method similar to the 2AFC, but as the name said, the observer's choice is forced between three alternatives Figure 4.4, of course only one is correct. This method is implemented by putting in two of the three choices the same stimulus (in this case sound) and in the other the different one that should be selected (the wrong one).

Was used the 2AFC method for the Initial Time Delay Gap (ITDG) and Reverberation Time (RT) tests and the 3AFC for the Density tests because the lasts are more difficult for the tester to understand what she/he has to select and with this method is easier to focus on which one is different.

Figure 4.4. Example of 3AFC in the website while listening to the sounds, just before making the choice.

4.3 Assumptions

The up-down method was used because of its high precision for the type of tests involved in this thesis, in fact this method is widely used for psychoacoustic tests.

It was decided to use the 1-up 1-down method that has the efficiency of estimating only 50% of the correct level shown by the results, but in combination with the three types of steps used for the first three turning points, the estimated level of the result is reached very quickly. Only after one more turning point it starts to calculate the average for the final result, in this way it has the arithmetic average of the last 8 turning points that begin, or at least should, from a level similar to the result that will be obtained.

In order to create the tests properly, was decided to set up a website, allowing people to do them in the simplest and most accessible way. Besides,the fact that you can also do these tests with your smartphone made the website more portable.

For the ITDG and Reverberation length tests, it was decided to use 4 different types of reverbs of length 0.5, 1, 2, 4 s (increasing exponentially) to understand if the human ear responds differently to these types of IRs. Regarding the density test, was only used one type of reverb, because the fundamental property on which the test is based is the Haas phenomenon Chapter 2.4, which has value only for the first 50 ms and that is the reason why doing more tests with different types of reverberation times would have been of little use.

Two types of sounds were chosen for each test:

- instrumental sample (sax) with frequency spectrum between 300 Hz and 10 kHz, 2.5 s long with fast attack and long decay, same for all the three tests
- vocal sample with frequency spectrum between 100 Hz and 3 kHz, 0.8 s long with medium attack and decay for ITDG and Reverberation length tests

- vocal sample with frequency spectrum between 100 Hz and 10 kHz, 3.1 s long with medium attack and decay for the Density test

The samples used in the tests were recorded in anechoic conditions so as not to influence the reverberations used in this research.

When the demo of the website was ready and the algorithm properly worked, it has been empirically found out how much the step has to be for each test, how many steps there are in total and the value of the big, medium and small step.

Type of test	Step	Max step
ITDG	5ms	200ms
Reverberation Time	50ms	2sec
Density	1ms	40ms

Table 4.2. Table of steps and the maximum step of tests.

By setting the maximum step in Table 4.2 there is a big difference between the reference sound and the one with the maximum step, this immediately helps the testers to recognize the right choice at the beginning of each test.

Quantity of steps	Big step	Medium step	Small step
40	10	3	1

Table 4.3. Value of steps and amount of those for all three types of tests.

4.4 Creation of Impulse Responses

To create the sounds to be used in the tests it has been used MATLAB, a program with high-performance language for technical computing, making the convolution between audio samples recorded in an anechoic chamber, with about zero reverb when it was recorded, and added reverbs, specially made.

To get all the audio needed for the tests it was decided to use MATLAB instead of doing all the algorithms in JavaScript because in this way it was were able to create them in a more correct, precise way, and the reverb parameters were professionally controlled.

4.4.1 Mirror image method

In geometrical room acoustics, the idea of a sound wave is replaced by the concept of a sound beam. This method is based on the principle that a specular reflection can be geometrically created by mirroring the source in the reflecting surface plane. It is used to construct the mirror image of a source.

In 1972, Gibbs and Jones [33] demonstrated that, using the imaging method for digitally calculating, the distribution of sound pressure level in a rectangular room produced more accurate results than with conventional methods. Thereafter, in 1978, Allen and Berkley [34] used the image method to perform impulse response (IR) calculations for box-shaped rooms with specular and angle independent reflections.

The IR, in the mirror image method, was created by modelling a talker in a room as a point source in a rectangular cavity. As said by the study [34], the procedure has been described in the following steps.

A single-frequency point source of acceleration in free space produces a pressure wave with the shape

$$P\left(\omega, \mathbf{X}, \mathbf{X}'\right) = \frac{\exp[i\omega(R/c - t)]}{4\pi R}, \tag{4.1}$$

where $P =$ pressure, $\omega = 2\pi f$, $f =$ frequency, $t =$ time, $R = |\mathbf{X} - \mathbf{X}'|$, $\mathbf{X} =$ vector talker location (x, y, z), $\mathbf{X}' =$ vector microphone location (x', y', z'), $i = \sqrt{-1}$, $c =$ speed of sound.

When a rigid wall is in place, the rigid boundary condition (zero normal velocity) can be achieved by symmetrically setting an image on the extreme side of the wall. Thereby,

$$P\left(\omega, \mathbf{X}, \mathbf{X}'\right) = \left[\frac{\exp\left[i(\omega/c)R_{+}\right]}{4\pi R_{+}} + \frac{\exp\left[i(\omega/c)_{R-}\right]}{4\pi R_{-}}\right] \exp(-i\omega t), \tag{4.2}$$

The two distances between the microphone and the source R_{-} and the image R_{+} are defined by

$$\begin{aligned} R_{-}^2 &= \left(x - x'\right)^2 + \left(y - y'\right)^2 + \left(z - z'\right)^2, \\ R_{+}^2 &= \left(x + x'\right)^2 + \left(y - y'\right)^2 + \left(z - z'\right)^2. \end{aligned} \tag{4.3}$$

The situation is more complicated in the general case of six walls since each image is itself imagined. The pressure can therefore be written

$$P\left(\omega, \mathbf{X}, \mathbf{X}'\right) = \sum_{p=1}^{8} \sum_{r=-\infty}^{\infty} \frac{\exp\left[i(\omega/c)\left|\mathbf{R}_p + \mathbf{R}_r\right|\right]}{4\pi\left|R_p + \mathbf{R}_r\right|} \exp(-i\omega t), \tag{4.4}$$

where R_p refers to the eight vectors given by the eight permutations and r is the integer vector triplet (n, l, m), and

$$\mathbf{R}_{\mathrm{r}} = 2\left(nL_{\mathrm{x}}, lL_y, mL_z\right), \tag{4.5}$$

where (L_x, L_y, L_z) are the dimensions of the room. Equation 4.4 is the pressure frequency response within rigid walls for a point source at $X = (x, y, z)$ and a receiver at $X' = (x', y', z')$. If Equation 4.4 is Fourier transformed, was obtain the room impulse response function

$$p\left(t, \mathbf{x}, \mathbf{X}'\right) = \sum_{\rho=1}^{8} \sum_{r=-\infty}^{\infty} \frac{\delta\left[t - \left(\left|\mathbf{R}_p + \mathbf{R}_r\right|/c\right)\right]}{4\pi\left|\mathbf{R}_p + \mathbf{R}_{\mathbf{r}}\right|}. \tag{4.6}$$

As mentioned earlier this method can only be used to recreate IR for box-shaped rooms so it is not valid for all the others, a limiting factor, which however in this study is not very relevant since the perception of three parameters of a generic reverberation will be evaluated.

4.4.2 Parameters for Impulse responses

The Room Impulse Response (RIR) is calculated by using a very powerful function in MATLAB, created by Stephen G. McGovern. This function is based on the use of the mirror image method [35], which is an improved version of the technique created by Allen and Berkley, but optimised to address the two main problems of redundant and unnecessary calculations. Therefore, these issues are solved by using look-up tables to avoid redundant calculations and the use of a sorting method to prevent unnecessary calculations. When used individually, these techniques provide a significant reduction in computation time, but the best result is obtained by using both of them.

Setting the parameters of the function to obtain the RIR:

- FS = sample rate of the RIR (used the same of the sample audio to convolve with)
- MIC = vector that contains the x, y, z coordinates of the microphone
- N = number of virtual sources
- R = reflection coefficient
- RM = row vector giving the dimensions of the room
- SRC = vector that contains the x, y, z coordinates of the sound source

4.4.3 Impulse responses for Initial Time Delay Gap tests

In order to create IRs for Initial Time Delay Gap (ITDG) tests, two main iterations have been used: the first one was needed to create the basic RIR used for the single test (example ITDG 0.5 s), while the second one nested was used to create, with the basic RIR, all the RIR with different steps for each test.

The sample of the IR's was imported and put in a vector. As the sample check has been done, if the sample was stereo, it merged into one vector (mono vector) and since the RIR is also mono, this passage is strictly needed to do the frequency convolution. Thereafter, all the parameters for the RIR were set, picked out empirically, except for the *RM*, which changes every iteration to obtain the reverb with 0.5, 1, 2, or 4 s. The RIR was created with the function described in Chapter 4.4.2. The next step was to continue setting how many zeros are in nanoseconds and how many steps to do for the test and how to have the step in ms, Table 4.2.

Creating the reverberated sample with the *freq_conv* function, to obtain a more realistic sound, the convolved signal was filtered with a lowpass with the frequency set as $(3/10) * FS$. After that, in order to reduce the loudness of the reverb and obtain a natural effect, the reverberated sample was added, dividing it by 2 in amplitude, to the original sample to obtain the complete audio. To not have a clipped signal, the complete audio ($s(n)$) was normalized by multiplying the complete audio $*0.9$ and dividing it by the max value (m_s) of the audio as in Equation (4.7).

$$m_s = \max(\text{abs}(s(n))), \tag{4.7a}$$

$$s_n(n) = 0.9 * \frac{s(n)}{m_s}, \qquad \forall n = 1, \ldots, N. \tag{4.7b}$$

In the end, the normalized audio ($s_n(n)$) was saved with the proper name in a *.wav* file.

In order to create the steps for the test, the number of zeros contained in a millisecond was set. How many steps to take for the test and how the step should be in ms is defined in Table 4.2.

After making the reference signal, the next step was to begin the iterations to create the audio with more ITDG. For the step of the ITDG, the number of zeros contained in a millisecond is multiplied by the step in ms defined in Table 4.2 and multiplied once again by the iteration index. The number obtained is the number of zero padding to be added to the beginning of the reference RIR created earlier. Thereafter, the reverberated sample was created with the *freq_conv* function and to obtain a more realistic sound, was filtered the convolved signal with a lowpass with the frequency set as $(3/10) * FS$. After that, the reverberated sample was added (divided by 2 in amplitude) to the original sample to obtain the complete audio, and to not have a clipped signal the complete audio was normalized by multiplying the complete audio $*0.9$ and dividing it by the max value of the audio as in Equation (4.7). In the end, the normalized audio was saved with the proper name in a *.wav* file.

This was done until the last step to be created was reached.

4.4.4 Impulse responses for Reverberation Time tests

For what concerns IRs for Reverberation Time (RT) tests, there are two main iterations: the first one used to create the basic RIR for the single test (example Reverb 4 s), and the second one nested used to create, with the basic RIR, all the RIR with different steps for each test.

In order to create the IR's for these tests, the sample was imported and put in a vector. As the control of the sample has been done, if it was stereo, it merged into one vector (mono vector), because the RIR is mono, which is strictly needed to do the frequency convolution. After setting all the parameters for the RIR, pick out empirically, except for the RM that changes every iteration to obtain the reverb with 0.5, 1, 2, or 4 s. The RIR was created with the function described in Chapter 4.4.2.

Thereafter, the reverberated sample was created with the *freq_conv* function and to obtain a more realistic sound the convolved signal was low-pass filtered with the frequency set to $(3/10) * FS$ and divided by 2, 1.8, 1.6, 1.5 (respectively to the different lengths) in amplitude, moreover, in order to reduce the loudness of the reverb and obtain a natural effect, the zero padding was added at the end of the signals to reach a similar length to the maximum of the test. Considering that it is necessary to avoid the problem of selecting the sound that is physically shorter in time, for these tests with the reverb of 0.5 s this padding is highly important. The next step was to continue adding the reverberated sample to the original sample to obtain the full audio, and in order to avoid having a clipped signal, the full audio ($s(n)$) was normalized by multiplying the full audio $*0.9$ and divided by the maximum value (m_s) of the audio as in Equation (4.7). Finally, the normalised audio ($s_n(n)$) was saved with the appropriate name in a *.wav* file.

The step to be taken in the RT tests is 0.05 s as written in Table 4.2 and the value to be added to the reverberation length was found empirically.

Having created the reference signal, continue by starting the iterations to create the audio with a higher RT at each step. The RIR had to be recalculated at each iteration with the dimensions of the reference room plus the step calculated in step before multiplied by the iteration index.

The reverberated sample was created with the *freq_conv* function and to obtain a more realistic sound, the convolved signal was low-pass filtered with the frequency set to $(3/10) * FS$ and divided by divided by 2, 1.8, 1.6, 1.5 (respectively to the different lengths) in amplitude, moreover, in order to reduce the loudness of the reverb and obtain a natural effect, the zero padding was added at the end of the signals to reach a similar length to the maximum of the test to avoid the problem of select the sound that is physically shorter in time, for this tests with the reverb of 0.5 s this padding is very important. After that, continued adding the reverberated sample to the original sample to obtain the full audio, and to avoid having a clipped signal the full audio was normalized by multiplying the full audio $*0.9$ and divided by the maximum value of the audio as in Equation (4.7). Finally, the normalised audio was saved with the appropriate name in a *.wav* file.

This was done until the last step to be created was reached.

4.4.5 Alternative methods for Density tests

Three other ways of doing the density test were tried, nevertheless it was found out that the best result obtained is with the IR's for the Density tests mentioned later (Chapter 4.4.6) which was chosen for several reasons. The first reason is that using an IR that simulates a natural reverb in a room, the space between the early reflections will be a few milliseconds and finite, so by adding other samples between them after not many iterations, you will have the saturation of samples without having zero paddings between them, so as to not have the possibility to have many steps to do the tests. Another reason is that the other methods create a high distortion in the frequency spectrum and phase after several iterations. To conclude, from a subjective point of view this method remains more natural to the ear, even after 40 iterations than the other three.

As explained in Chapter 2.4, the most important part of the sound, to recognize if a sound is present, are more or less the first 50 ms of the sound. In Figure 4.5 are shown the first 50 ms of the IR used for these tests, so according to this theory it took a reverb with 2 s of reverberation time and it was tried to modify it in different ways in order to obtain a density effect, evaluating by checking frequency and phase the reverbs and checking if they are suitable to have enough intervals to be able to take the test.

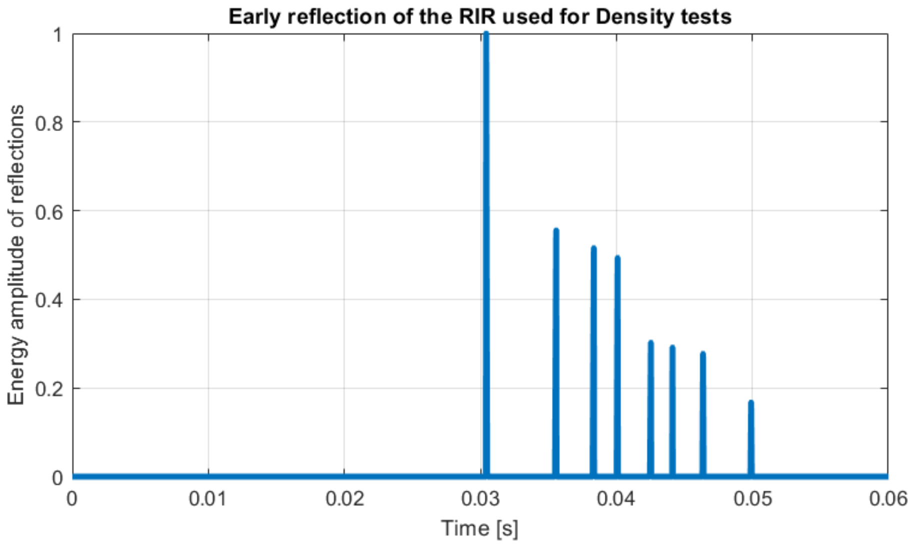


Figure 4.5. Early reflections of basic reverb used in Density tests.

First alternative method

In the first method it was tried to add a sample between two samples in every iteration, with the amplitude equal to the mean of the two samples Figure 4.6, but two problems occurred:

- Boost of some frequencies, like notch filters;
- After 7-8 iterations there was no zero between some samples (depending on the Sample frequency) so it was useless to continue with the iterations.

Therefore, for both problems, this approach is not appropriate for this test.

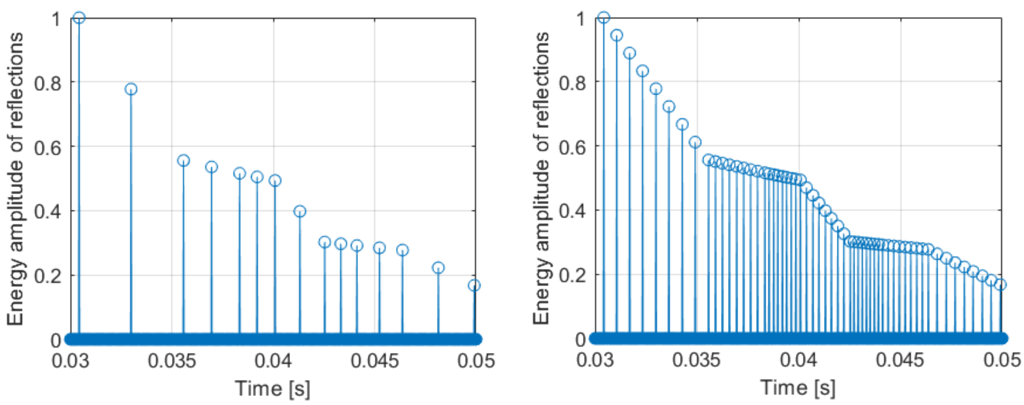


Figure 4.6. Early reflections of first alternative method for density test, first iteration (left), third iteration (right).

Second alternative method

In the second method, it was tried to add a repetition of every sample at a well-defined distance in every iteration Figure 4.7, but two problems occurred:

- Boost of some frequencies, like bell parametric shape;
- After 10-11 iterations some of the repetitions overwrite the samples that follow, so it is useless to continue with the iterations.

For both problems, this approach is not appropriate for this test.

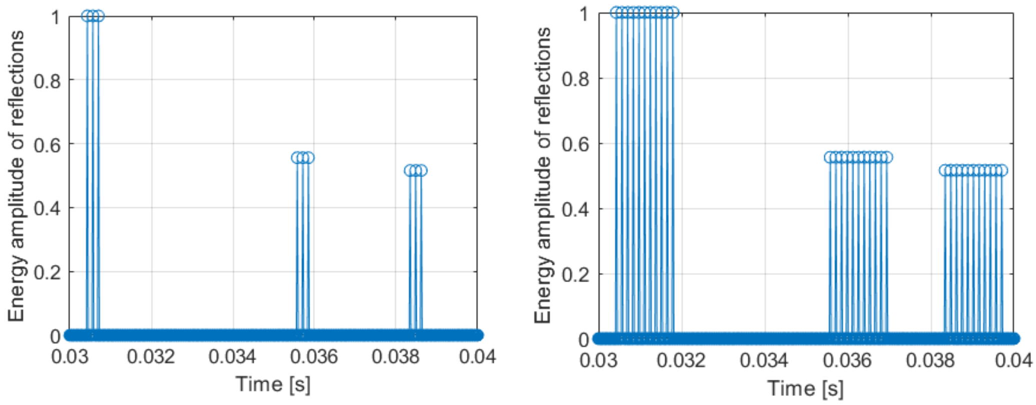


Figure 4.7. Early reflections of second alternative method for density test, second iteration (left), tenth iteration (right).

Third alternative method

To solve the problems of the second method, it was attempted to mix with the method used in Chapter 4.4.6 adding zero padding after the repetitions created equal to the distance between them Figure 4.8.

Using this method, the problem of having more than around 10 steps is solved, but the frequency response of these reverbs contains an enhancement of frequency, which is less than the previous one, but still has also changed the distance between the repetitions.

Method chosen for Density tests

After several trials and changes, the method used for the density test was done by adding zero padding between the samples of the first 50 ms of the RIR using the step in ms as shown in the Table 4.2.

This result was achieved because this method does not have the problems encountered with the other three methods. The frequency spectrum and also the phase remained unchanged, furthermore, you can add as many zeros as you want, which theoretically creates infinite steps but having, from the auditory point of view, a completely distorted signal, that is because beyond a certain threshold you would get a sort of delay.

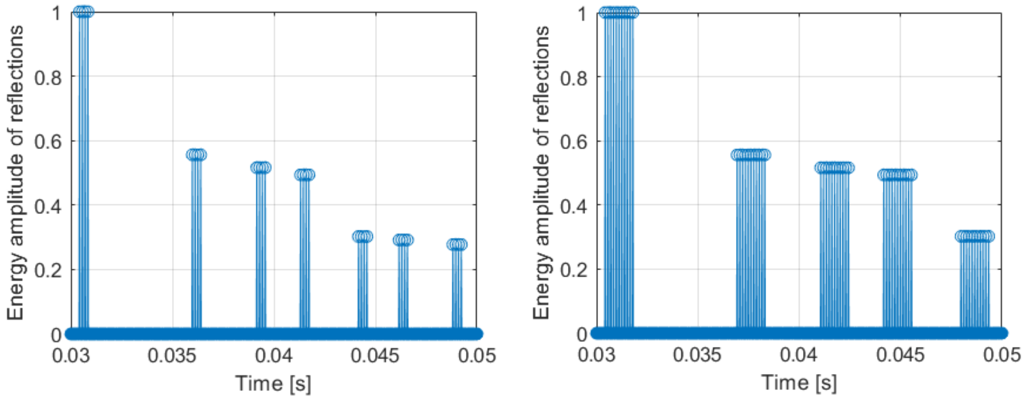


Figure 4.8. Early reflections of third alternative method for density test, third iteration (left), tenth iteration (right).

4.4.6 Impulse responses for Density tests

There is only one iteration which is used to create, with the basic RIR, all the RIR with different steps for both tests.

The sound sample was imported and put into a vector. As the sample check has been done, if the sample was stereo, it merged into a vector (mono vector), it is strictly necessary to do the frequency convolution because the RIR is too. Then, all the parameters for the RIR were set empirically constant to obtain reverberations with RT (the factor 2 split explained two sentences later is also taken into account) 2 s. The RIR was created with the function described in Chapter 4.4.2.

The reverberated sample was created with the *freq_conv* function and to obtain a more realistic sound the convolved signal was low-pass filtered with the frequency set to $(2/10) * FS$ and divided by 2 in amplitude in order to reduce the loudness of the reverb and obtain a natural effect. After that, it continued adding the reverberated sample to the original sample to obtain the full audio and to avoid having a clipped signal the full audio was normalized by multiplying the full audio $*0.9$ and divided by the maximum value of the audio as in Equation (4.7). Finally, the normalised audio was saved with the appropriate name in a *.wav* file.

After setting up how many steps you wanna create for these tests, the step was set with the quantity of zero padding to be inserted between the first 50 ms non-zero samples. In reference to this, see the Table 4.2.

After finding the first reflection (ITDG), the iterations were started in order to create the audios by adding zero padding between the first 50 ms of samples, practically lowers the density of the early reflections.

The next step was creating the reverberated sample with the *freq_conv* function and to obtain a more realistic sound the convolved signal was low-pass filtered with the frequency set to $(2/10) * FS$ and divided by 2 in amplitude so as to reduce the loudness of the reverb and obtain a natural effect. After that, it continued adding the reverberated sample to the original sample to obtain the full audio, to avoid having a clipped signal the full audio was normalized by multiplying the full audio $*0.9$ and divided by the maximum value of the audio as in Equation (4.7). Finally, the normalised audio is saved with the appropriate name in a *.wav* file.

This was done until the last step to be created was reached.

4.5 Online test design

Due to the current pandemic and the restrictions in place, it was not possible to carry out the tests in the laboratory, therefore it was decided to create a website.

Talking about the creation of the website, the keywords were Simplicity and Accessibility. The most important concept of the website was thinking about being simple to use with a simple interface, all the basic information on the homepage, and the explanation of tests, step by step, before doing them. Thanks to the responsiveness of the website and the possibility to choose three different languages, everyone can simply access the website with his smartphones, tablets, or computers.

In order to create the website, these languages were used:

- HTML (HyperText Markup Language): a markup language for formatting and pagination of hypertext documents;
- CSS (Cascading Style Sheets): a language used to define the formatting of documents, in this case HTML;
- JS (JavaScript): an object-oriented programming language used for front-end and back-end.

While, to create the website the following features were used:

- Brackets: a source code editor with a primary focus on web development;
- Bootstrap: a free and open-source CSS framework directed at responsive, mobile-first front-end web development;
- PostgreSQL: also known as Postgres, is a free and open-source relational database management system (RDBMS) emphasizing extensibility and SQL compliance;
- Swagger: an Interface Description Language for describing RESTful APIs expressed using JSON;
- Node JS: an open-source, cross-platform, back-end JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser;
- Knex.js: a "batteries included" SQL query builder for Postgres designed to be flexible, portable, and fun to use.

4.5.1 Front-end

The website is based on 6 pages:

- The homepage that contains all the link to reach the three typologies of tests, names, and GIFs to see graphically which type of test you are going to select, basic information that the user has to know to do the tests

- Tests' selection page that contains the information to understand what to do in that type of tests, 2 buttons with associated sounds to better understand what to evaluate and a menù to select which test do
- Tests' page that contains the practical information to do the test, the play button to listen to the sounds, 2 (ITDG and Reverberation time tests) or 3 (Density tests) buttons for the selection of the sound, the progression bar, and the finish button to conclude the test (that became selectable only when the tester finish the test)
- Form page contains the result of the test done, all the information that the tester has to select for the database Chapter 4.5.3, and the send the result button
- Redirect page that contains 4 buttons, one for redirect to the homepage and the other 3 are for redirect to the main typology of tests
- Contact me page that contains basic information of the creators with their links, information of the program used to create the website, and the special word of thanks to the people that help to improve it

All the pages have:

- header with the name of the website with the link for the homepage, change language menù to select a language between English, Italian and Polish, and the link for the contact me page
- footer containing the link for the contact page and the link of GitHub profile

The pages for the selection, the test, and the form have also the breadcrumbs bar to navigate more simply on the website.

4.5.2 Algorithm of the tests

According to the Table 4.3 has been used 40 different steps for every test with 3 different types of steps, big, medium, and small respectively 10 steps, 3 steps, and 1 step. The test will start with the big step set and after one fails or reaches the last step it becomes the medium and after the small one as you can see in the example in Figure 4.9.

The reference sound will not change during the test, only the non-correct one will start from the highest differences.

After the user clicks on the play button she/he will hear 2 (2AFC method) or 3 (3AFC method) sounds one after the other starting with the one with more difference, 200 ms more for the ITDG test, 2 s more of RT for the RT one and 40 ms of zero padding for the Density one.

The correct sound is the one for which in the condition of the 1up-1down method you go down and the incorrect sound is the one that makes you go up.

Thanks to the random function, a random number decides which sound has to be the reference one, the sound with the lowest ITDG or RT in the case of the homonymous tests is also the correct one on the contrary for the Density test the

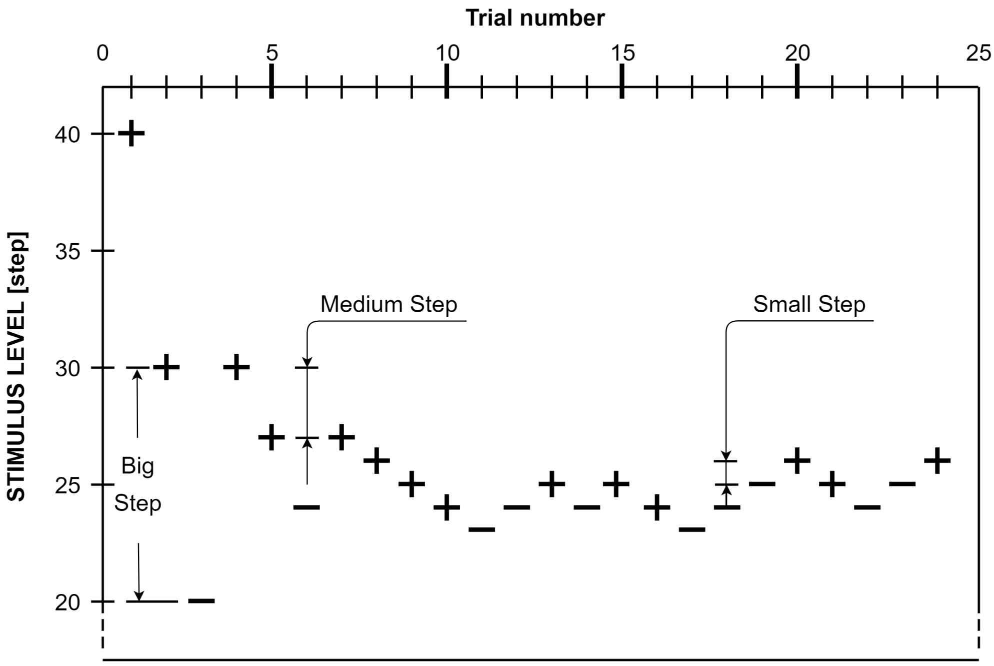


Figure 4.9. Example of a real 1-up 1-down test with the three displayed steps.

correct sound is the one with the lowest density i.e. the one with the highest zero padding and the two reference ones are the incorrect ones.

The user had to decide which sound is the correct one by clicking on the button:

- If the response is correct the button pressed will become green and the other/s one red, depending on the step (big, medium, or small) next time the non-correct sound will play with shorter ITDG, shorter Reverberation Length, or higher Density.

- If the response is non-correct the button pressed will become red and the other/s one green, depending on the step (big, medium, or small) next time the non-correct sound will play with higher ITDG, higher Reverberation Length, or lower Density.

The previous part will be repeated until the twelfth turning point is reached. At that point, the user has to click the finish button test and the *endingFunction()* will calculate the average of the last 8 turning points values.

4.5.3 Database (Back-end)

The database was made with node JS working on Postgres Database, for this website, one *post* operation was enough to insert eleven attributes in each row.

It was thought that for this type of tests ask users to send, in addition to the test result and the type of test that are automatic, also other basic information about themself and the characteristic of the device and the room in which they did the test to be more precise in drawing conclusions:

test id	type of test	result	hearing problems	age	gender	audio system
32	ITDG Test 1 second Sax sample	12.5	Yes, I have it now	50/60	female	In Ear headphones
33	Reverberation Time Test 0.5 second Polish sample	0.05625	Yes, I had in the past	40/50	male	Normal Audio System
34	Density Test 2 seconds Sax sample	3	No, never	20/30	female	Overhead Headphones

background noise	reverb knowledge	timestamp	person name	person surname	person email
Normal Background Noise	No	2021-02-25 18:48:29	NULL	NULL	email@example.com
Very Quite Environment	No	2021-02-26 09:34:42	NULL	NULL	NULL
High Background Noise	Yes	2021-02-26 12:25:57	Mario	Rossi	mario.rossi@example.com

Table 4.4. Example of three rows of data received in the DataBase.

- age
- gender
- hearing problems
- audio system
- noisy environment
- acoustic knowledge

The table *test* contains 11 columns in total as is shown in Table 4.4:

- *test_id*: the primary key of the table, type integer auto-generated in the database, used it to take into account the order of results' arrival
- *type_of_test*: type string, containing the name of the test done

  - ITDG Test 0.5 (1, 2 or 4) second Polish (or Sax) sample
  - ReverberationTime Test 0.5 (1, 2 or 4) second Polish (or Sax) sample
  - Density Test 2 seconds Polish (or Sax) sample

- *result*: type real (number), containing the associated test result
- *hearing_problems*: type string, containing one of the three replies to the question "Do you have any hearing problems?"
  - Yes, I have them now
  - Yes, I have had them in the past
  - No, never
- *age*: type string, containing the age range of the tester
  - -20 (20 or less)
  - 20/30
  - 30/40
  - 40/50
  - 50/60
  - 60+ (60 or more)
- *gender*: type string, to know the tester's gender
- *audio_system*: type string, containing the typology of audio system used for the test
- *background_noise*: type string, containing low-level information about the environment in which the tester did the test
  - High background Noise
  - Normal background Noise
  - Very quiet environment
- *reverb_knowledge*: type string, to know if the tester knows scientifically reverbs (yes or not)
- *timestamp*: type timestamp without time zone, generated by the JavaScript while doing the post

And other 3 non-mandatory fields that could be NULL:

- *person_name*: type string, name of the tester
- *person_surname*: type string, the surname of the tester
- *person_email*: type string, email of the tester in which all the results will be sent

4.6 Participants

As mentioned before, these tests were created so that anyone could take them comfortably with their own device. Participants can be of any age or gender and can take the test in any environmental conditions. Obviously, they were advised on the homepage of the site what the best conditions are to do these tests so that they can get more accurate results and they reported that they do not have severe hearing problems. On the website in the Front-End Chapter 4.5.1 the participants were given an exhaustive explanation of the type of test to be done with a scientific clarification of the parameter they would be testing, an informal explanation of it and finally also two buttons containing reference sounds at the extremes of the range to be tested so that they could also have an auditory comparison.

4.7 Reviews after demo test

The advice received after the first demo was tried by colleagues was welcomed. Before the 1-up 2-down method was implemented, but after the first testers tried the tests they thought that it was too long with 14 turning points, so to speed-up the tests, it was changed with the 1-up 1-down method with 12 turning points for all the tests also if it is not the most precise. The final value is made by the average of the values of the last 8 turning points.

When the website was complete it was decided, before sharing the link to everyone, to ask some other colleagues and friends to try some tests and find any bugs or problems with the tests. Everything worked well, some of them have advised making it more understandable even to a less experienced audience and so it was.

Chapter 5

Test results and discussion

Thanks to the various people who took the tests, there were 382 results in total, 47 for the two density tests, 172 for the 8 Initial Time Delay Gap (ITDG) tests, and 163 for the remaining 8 Reverberation Time (RT) tests.

5.1 Main Results

Six columns are included in the tables below to summarise all the main results of the study:

- the type of signal used, vocal or instrumental
- the duration of reverberation applied to the signal in seconds
- the arithmetic mean of the results, in milliseconds or seconds depending on the type of test
- the Standard Deviation (SD) indicating how much the results vary around the mean
- skewness, which indicates how non-symmetrical a normal distribution is with respect to the mean in Figure 5.1; if positive, it is called right skew, if negative, left skew.
- the median of all results, i.e. the central result by having all results in ascending order.

5.1.1 Initial Time Delay Gap test results

In this section there is Table 5.1 for the results of ITDG that contain 9 rows because in addition to the 8 tests made, the one with the union of all of them was added because the ITDG does not depend strongly on the length of the reverberation. Following a graph in Figure 5.2 with mean and SD of the ITDG test results for both sound samples at varying RT.

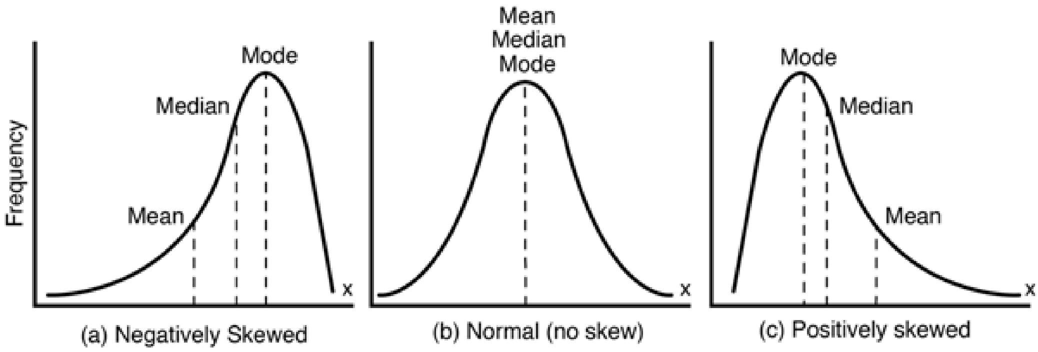


Figure 5.1. Symmetric and asymmetric Gaussian distributions. [36]

Initial Time Delay Gap					
type of sample	RT [s]	mean [ms]	standard deviation	skewness	median [ms]
vocal instrumental	0.5	57.7 55.4	68.5 47.2	0.96 1.02	14.4 28.8
vocal instrumental	1	36.1 32.9	50.3 23.8	1.95 1.36	12.5 25.0
vocal instrumental	2	32.1 66.0	54.4 55.3	2.23 0.76	10.0 41.3
vocal instrumental	4	61.8 65.2	68.4 66.2	0.83 1.07	14.4 32.5
vocal and instrumental	all	51.4	57.1	1.30	21.9

Table 5.1. Table that contain all the results of ITDG tests.

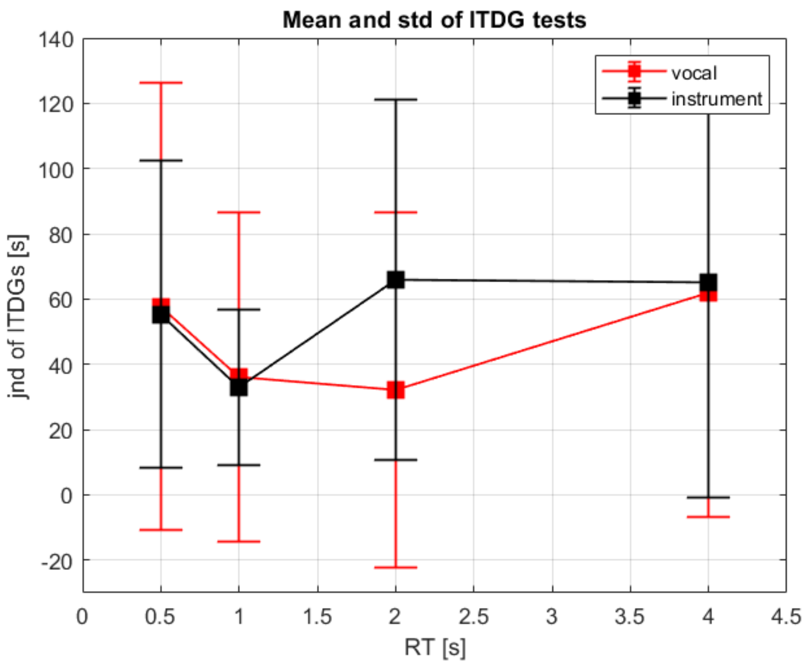


Figure 5.2. Mean of ITDG tests with SD at variation of RT for both samples.

5.1.2 Reverberation Time test results

In this section there is Table 5.2 for the results of RT that contain 8 rows one for each of the tests and a column was added with the average as a percentage of reverberation length because it was found to be strongly correlated. Following a graph in Figure 5.3 with mean and SD of the RT test results for both sound samples at varying RT.

Reverberation Time						
type of sample	RT [s]	mean [s]	standard deviation	skewness	median [s]	mean [%]
vocal	0.5	0.256	0.434	2.56	0.091	51.2
instrumental		0.195	0.281	1.99	0.081	39.0
vocal	1	0.360	0.437	1.42	0.113	36.0
instrumental		0.325	0.510	2.10	0.100	32.5
vocal	2	0.885	0.752	0.18	1.000	44.3
instrumental		0.662	0.582	0.51	0.700	33.1
vocal	4	1.411	0.655	-0.98	1.725	35.3
instrumental		1.473	0.507	-1.36	1.575	36.8

Table 5.2. Table that contain all the results of RT tests.

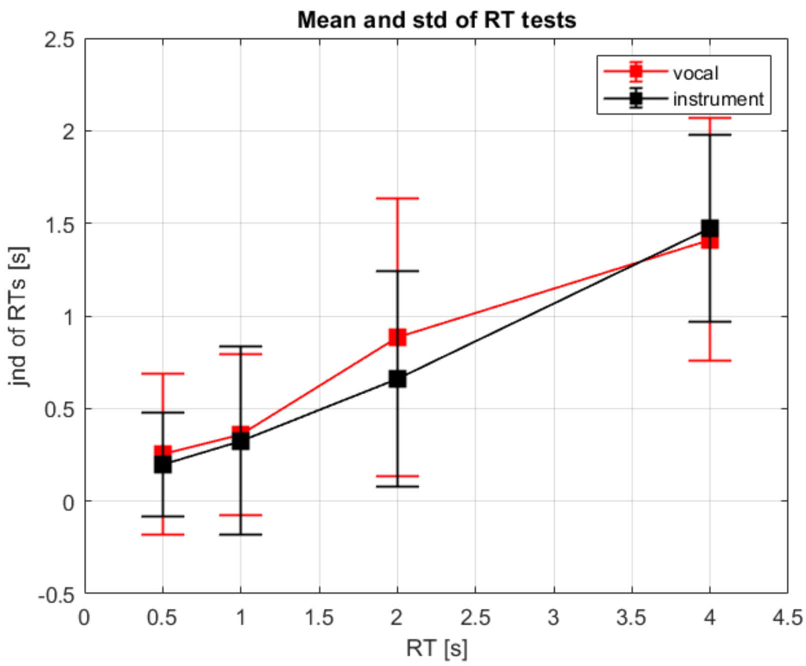


Figure 5.3. Mean of RT tests with SD at variation of RT for both samples.

5.1.3 Density test results

In this section there is Table 5.3 for the results of Density that contain 3 rows containing the two tests plus the total formed by the union of the two. This is followed by a graph in Figure 5.4 with the mean and SD of the Density test results for both sound samples in the left position, in the middle the results filtered to have only those from testers with no hearing problems and on the right the results filtered from those with no acoustic knowledge.

Density					
type of sample	RT [s]	mean [ms]	standard deviation	skewness	median [ms]
vocal instrumental	2	19.2 15.5	14.3 13.2	0.08 0.52	21.3 12.8
vocal and instrumental	all	17.2	13.7	0.31	15.3

Table 5.3. Table that contain all the results of Density tests.

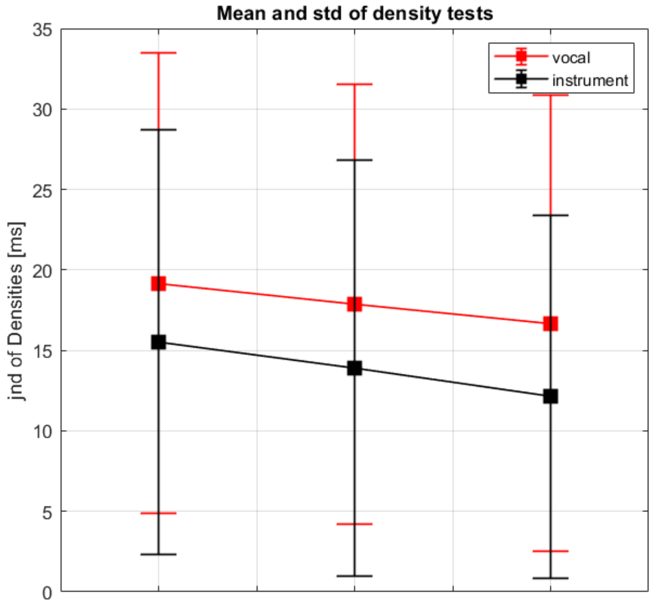


Figure 5.4. Mean of density tests with SD variance for both samples of total results (left), results without hearing problems (middle), and results of those with acoustic knowledge.

5.2 Statistical Analysis

For this study, the 'classical' parametric method was initially used, but the results were inconclusive because having not normally distributed data did not work. In order to find reliable results, two Bayesian analyses were used which are based on Bayes' theorem, in which the probability is calculated and updated after obtaining new data. The first method is Bayesian Linear Regression that performed a robust analysis for non-normal distributions, this analysis provides BF factor values to understand how much the data is influenced by one of the parameters, in this case, how much the RT or sample type influences the perception of the three parameters analysed in this thesis. The only shortcoming of this method is that one does not have full control over the a priori assumptions, so a second analysis was done using the JAGS module based on Markov Chain Monte Carlo.

Markov Chain-based Monte Carlo methods (MCMC) are a class of algorithms for sampling from probability distributions based on the construction of a Markov chain having the desired distribution as the equilibrium (or stationary) distribution. After simulating a large number of steps in the chain one can then use the extracted values as a sample of the desired distribution. Usually, it is not difficult to construct a Markov chain with the desired properties, but it is not always possible to determine a priori how many steps are necessary to converge with an acceptable error to the stationary distribution.

With the MCMC method, you have to provide the whole data equation, including the a priori assumption of the coefficients, in this case, you can control all possible interactions and even change the a priori assumptions to check if the result changes. The results were stable even with different distributions and different ranges of the coefficients.

The MCMC method was set with the *intercept* coefficient for the results obtained, the *beta2* coefficient with the type of sample used for the test (vocal or instrumental) and only for the ITDG and RT test also the type of RT used with the *beta1* coefficient.

5.2.1 Statistical analysis of Initial Time Delay Gap tests

In Table 5.4 are the most important results (derived from Appendix B) of the statistical analysis done with the MCMC method which reveals a tendency for the ITDG to be perceived at around 44 ms, without strong correlations with respect to the type of sample used or the duration of the reverb.

		Posterior		95% Credible Interval	
Parameter	**Mean**	**Median**	**SD**	**Lower**	**Upper**
intercept	44.054	43.918	11.527	21.366	66.727
beta1	4.636	4.613	2.998	-1.126	10.609
beta2	-1.340	-1.390	6.281	-13.705	10.978

Table 5.4. Markov chain Monte Carlo method summary of ITDG tests.

5.2.2 Statistical analysis of Reverberation Time tests

In Table 5.5 are the most important results (obtained from Appendix B) of the statistical analysis made with the MCMC method for the RT tests which reveals a tendency to be perceived differently according to the different length of the reverberation thanks to the mean value of *beta1* which is high compared to the mean result of *intercept*, regarding the type of sample used this method detects a slight correlation, but negligible.

	Posterior			95% Credible Interval	
Parameter	**Mean**	**Median**	**SD**	**Lower**	**Upper**
intercept	-0.066	-0.065	0.146	-0.354	0.222
beta1	0.355	0.355	0.031	0.295	0.417
beta2	0.063	0.062	0.085	-0.103	0.231

Table 5.5. Markov chain Monte Carlo method summary of RT tests.

5.2.3 Statistical analysis of Density tests

In Table 5.6 are the most important results (derived from Appendix B) of the statistical analysis done with the MCMC method for the Density tests which denotes a slight correlation with the type of sample used thanks to the *beta2* coefficient and a Just Noticeable Difference (JND) on average of 12 ms and SD 5.8.

	Posterior			95% Credible Interval	
Parameter	**Mean**	**Median**	**SD**	**Lower**	**Upper**
intercept	12.411	12.417	5.792	0.916	23.785
beta2	3.255	3.257	3.699	0.916	23.785

Table 5.6. Markov chain Monte Carlo method summary of density tests.

5.3 Discussion of Initial Time Delay Gap results

Thanks to the averages of the results it is possible to understand only approximately what could be the perceived JND because the standard deviation in all the results is very large and this can be seen very well looking at the histograms. For example, in Figure A.1 17 testers have obtained a result lower than 30 ms, but 5 have obtained results over 160 ms obtaining so an average of 57.7 ms and a standard deviation equal to 68.5 that is very big, in the other 7 tests of ITDG in average has been a little lower the Standard Deviation (SD). Even if the SD is high it can still be said that the perception of the ITDG does not depend strongly on the RT and this is also confirmed by the statistical analysis.

From the histograms proposed in Appendix A it is possible to state, differently from the results of the average and the statistical analysis, that in all 4 tests of the ITDG made with the vocal sample they obtained a result with the asymmetrical distribution mode lower than the instrumental one. Being that the results of the ITDG are strongly uncorrelated by varying the RT the vocal results for the 4 tests were summed and the same was done for the instrumental ones between them thus obtaining at Figure 5.5 with which it is more evidently noted that with the vocal sample lower results were obtained.

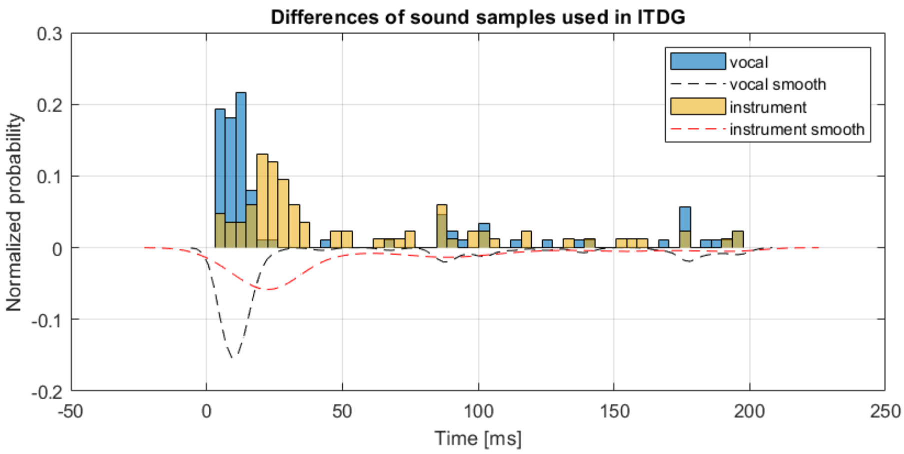


Figure 5.5. ITDG tests summing the vocal results by comparing them with the instrumental sum and a smooth version of them inverted with respect to the time axis.

Statistical analysis using the MCMC method indicates that the type of sample (vocal or instrumental) does not influence the test results of the ITDG, yet from the Figure 5.5 it would appear that it does, albeit slightly.

5.4 Discussion of Reverberation Time results

Regarding the Reverberation Time (RT) the results are a bit different than the ITDG, in this case, the perception of RT depends on the variation of the stimulus, the difference between the values of the test with 0.5 s of RT and the one with 4 s is huge both for mean and skewness and this is very clear thanks to histograms in Appendix A.

From the three different types of results, it can be seen that by changing the type of sample from vocal or instrumental, the RT is perceived without differences, so it can said that it is not dependent on the type of stimulus applied.

5.5 Discussion of Density results

The results for the density tests showed that with an instrumental sample is 15.50 ms, it is easier to recognize a reverb with higher early reflection density than speech is 19.16 ms with about is 4 ms difference in the mean however, this is only a hypothesis because as you can see in Table 5.3 the SD is high so the data is not accurate, however the statistical analysis with the MCMC method detects the correlation between the type of sample used and the perceived density value, which is not high, but present as it can seen in Table 5.6.

5.6 Other considerations

Since the way the tests had been conducted, unfortunately there are just a few people's results which are complete (which means they have done all the eighteen tests). That is because, for matters of privacy, not everyone put their name or surname on the test (given that it was an additional request) and it has made not possible to lead back to every tester. Another reason is that many others probably did not complete all the tests, but they partially done them which is the reason why it was not possible to draw conclusions to the single person.

5.6.1 Gender and age

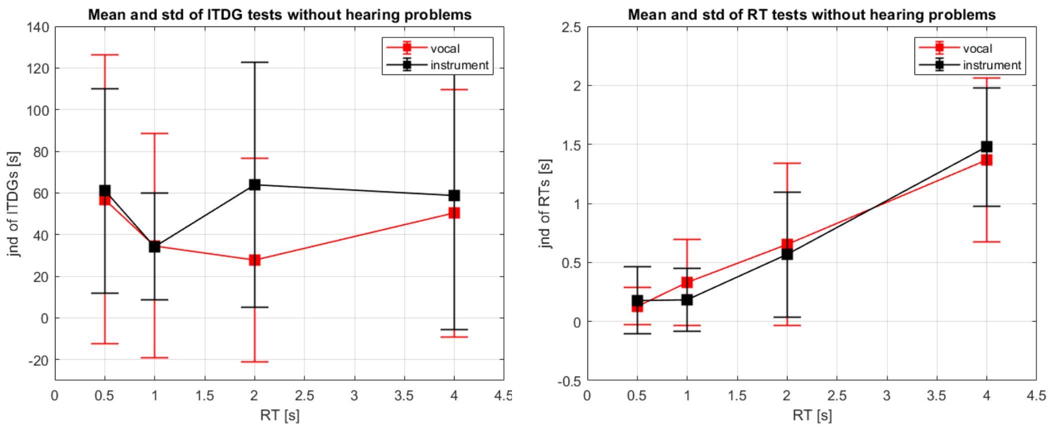
Concerning the results for gender, even if women who took the test were less than 32%, it has been tried to compare the results between women and men as the RT varied, but differences of any importance had been found.

As far as the age category is concerned, 284 out of 382 tests carried out, almost 75%, were done by young people aged between 20 and 30 years. Also in this case it is useless to make considerations because the percentage of the other groups in this category is much lower.

5.6.2 Hearing problems

81 of the 382 tests were done by people who reported having minor hearing problems or had hearing problems if these tests are not taken into account. As can be seen from Appendix C, the results are slightly more accurate having a smaller mean and SD mean for the ITDG tests and a clear reduction in percentage ($\sim 8\%$) for the RT tests.

Thanks to Figure 5.6a for the ITDG tests and Figure 5.6b for the RT tests, it can be seen that the trend is the same using all the data, i.e. ITDG tests constant as RT varies and RT tests that are linearly dependent on RT.

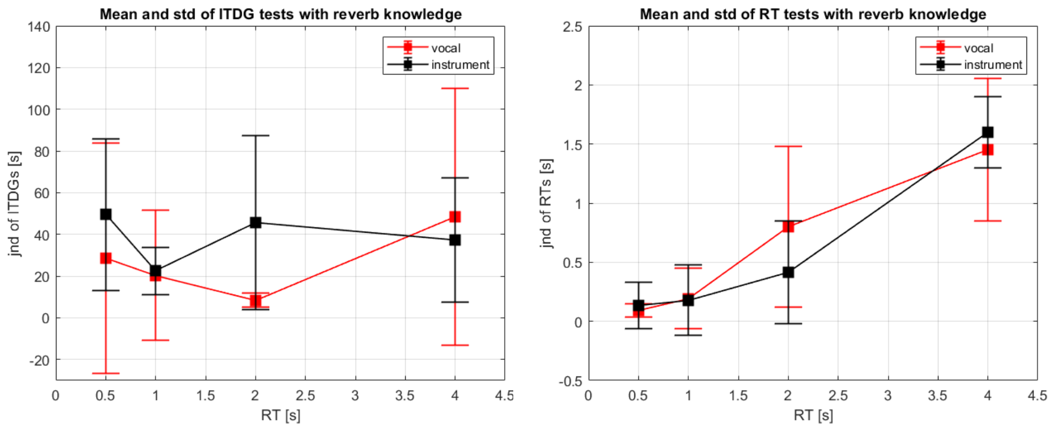


(a) ITDG (b) RT

Figure 5.6. Mean of ITDG and RT tests with SD at variation of RT for both samples just using the results with testers without hearing problems.

5.6.3 Acoustic knowledge

47% of the tests were done by people who claim to have no knowledge of acoustics and particularly reverberation. The main results using only the results from people with knowledge of acoustics are shown in Appendix C, which reveal that for the ITDG tests a much lower overall mean is achieved than with all results (sim20ms) and also a SD mean slightly more than half.



(a) ITDG (b) RT

Figure 5.7. Mean of ITDG and RT tests with SD a variation of RT for both samples using only results with acoustic knowledge testers.

Due to Figure 5.7a for ITDG tests and Figure 5.7b for RT tests, it can be seen that the tendency is the same as unfiltered results, i.e., ITDG tests constant as RT varies and RT tests linearly dependent on RT.

5.6.4 Audio system and noisy environment

As regards running the test, the majority of people who done it, that is the 81.68% (312 tests / 382 total tests) of the testers, used earphones or headphones. Since this is the best audio system to use, it means that the results are 80% reliable for what concerns this aspect.

Slightly less than 50% of the tests were done in a normal background noise environment, the rest of the tests were done in a very quiet environment except for one test that was done in high background noise conditions.

Data received using cell phone audio with high background noise were not considered because they were probably faulty, fortunately, only one result was done this way and that is from the voice density test with a value of 39.5 out of 40 which is a very bad result.

5.6.5 Final discussion

The JND of ITDG is 44.40 ms – 12 ms, which is the result of the statistical analysis with a deviation equal to the difference from the results obtained using only testers with acoustic knowledge Table C.1. The resulting JND obtained in this study is very different from those found in other studies that are more or less 10 ms, but in those cases the optimal conditions for the testers were applied, which is not possible in this study.

The general JND obtained from this study for RTs is $38,4\% - 11\%$ depending on the duration of the reverb and is formed by averaging the percentages over the type of RT used for the test. In the literature, the JND for RT is 5% as far as ISO [23] rules are concerned, but in another paper [21] with tests similar to the one done in this study here the JND is equal to 24.5% which is not very far from the one found.

The general value of JND derived from this study for the Density tests is $12.4\ \mathrm{ms} \pm 2\ \mathrm{ms}$ was created from the mean value found by statistical analysis with a deviation equal to half the difference between the two sample types used.

In Figure 5.8 the definition value D_{50} for the impulse response used in the Density test is plotted and due to the general result found earlier it can be argued that the JND is 0.35, which is about seven times higher than that described in the literature in Chapter 2.8. Not being the parameter to describe this type of test it was not expected to be very similar, but not so high either.

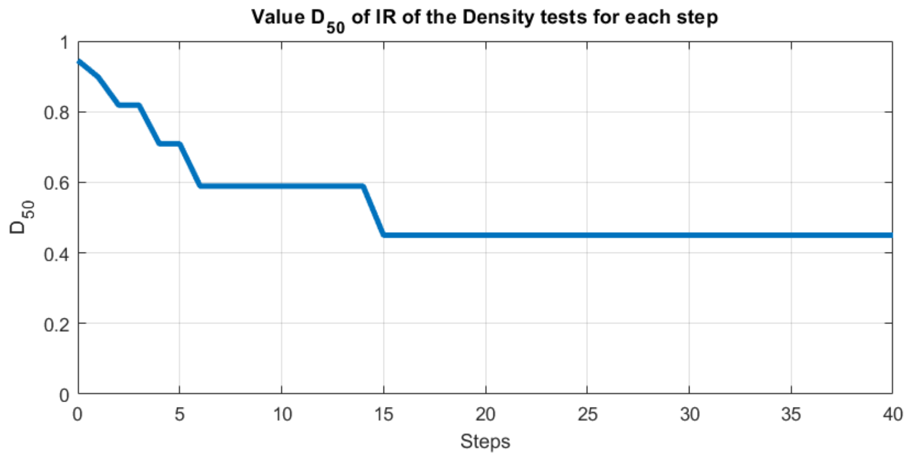


Figure 5.8. Plot of the D_{50} for each step of the impulse response used in Density tests.

Chapter 6

Conclusions and Future Works

In this study, psychoacoustic tests were carried out to determine the Just Noticeable Difference (JND) of three reverberation parameters, the Initial Time Delay Gap (ITDG), Reverberation Time (RT) and the density of early reflections. ITDG and RT were examined by varying the reverberation time while the perception of all three parameters was assessed by changing the type of sound sample employed.

From the test results, it can be stated that the JND for the RT tests increases with increasing reverberation time, while for the ITDG tests this incidence of JNDs was not found as can be seen in Figure 5.7, however the standard deviation for this parameter is very high which does not allow us to give more detailed conclusions. The results, varying the type of sample used, reveal a slight difference in the perception of the JND in the ITDG tests as shown in Figure 5.5, but since the standard deviation is high, it is not possible to give precise results on how much this difference is. Thanks to Figure 5.4 which presents the results of the Density tests it is possible to appreciate the difference in perception from the type of sound sample applied. The findings of the RT tests do not detect considerable differences from the type of sample used.

Analyzing the results only of the testers who did the test in an environment with normal or low ambient noise, with earphones or headphones, declared to have acoustic knowledge and no hearing problems, can be said to have obtained the result with lower mean and standard deviation, so more uniform and precise. If one could have had people with musical experience do the tests in a controlled environment, the final results would have been more accurate.

- The results, in general, are in line with the psychoacoustic findings, i.e., they reflect an asymmetric distribution with positive skewness (except for the two RT tests having reverberated samples with RT 4 seconds Figure A.9) although some results tend to be close to the maximum step, purposely chosen to start the test with a large difference in the first decision.

- Many testers, not experts in acoustics after repeating the same tests one or more times gave the feedback that they were able to improve their score.

Looking at the histograms one may state that in the tests with the total results of ITDG and Density (tests not correlated with the variation of RT) that there is a high peak and two others, one towards the middle of the range and another towards the end. This happens because the high peak is created by the results produced by the testers knowing what to choose and the average of it should be the real JND. The second peak (the one towards the center) is given by those who started the test not knowing perfectly what to answer and after a few wrong answers they realize the errors and improve, this phenomenon is caused by the type of method used for testing, the 1-up 1-down method, because making a mistake in one of the first answers the step decreases and it is difficult to get to a low value of the range in consideration, so in the future, the procedure could be modified so that the error on the first answer could be ignored, this might decrease the SD and then the results could be analysed in detail. Finally, is found the peak (towards the bottom) that represents those who did the test without knowing what they had to answer, it can argue this because the first sounds that are heard are very different from each other so if you select the wrong one you do not know the target of the test.

The methods explained in this thesis were chosen because of the COVID-19 situation and for the same reason, it was impossible to have homogeneous groups of testers.

In the future, in order to complete this study, tests could be carried out to calculate the JND for other characteristics of room reverberations by using only people with experience of listening to music and using ideal listening conditions to have more uniform and precise results.

Appendix A

Frequency Histograms

Frequency histograms are useful tools for understanding the data received from numerous tests. In this study they were used to graphically identify how the data were distributed, find outliers, and understand trends in the results.

ITDG

In this section there are 4 histograms for the different reverbs used for the ITDG test and in each one there are the two types of samples used differentiated by colour.

At the end, a histogram with the sum of all results obtained in all 8 tests was added, due to the fact that ITDG is not strongly dependent on RT.

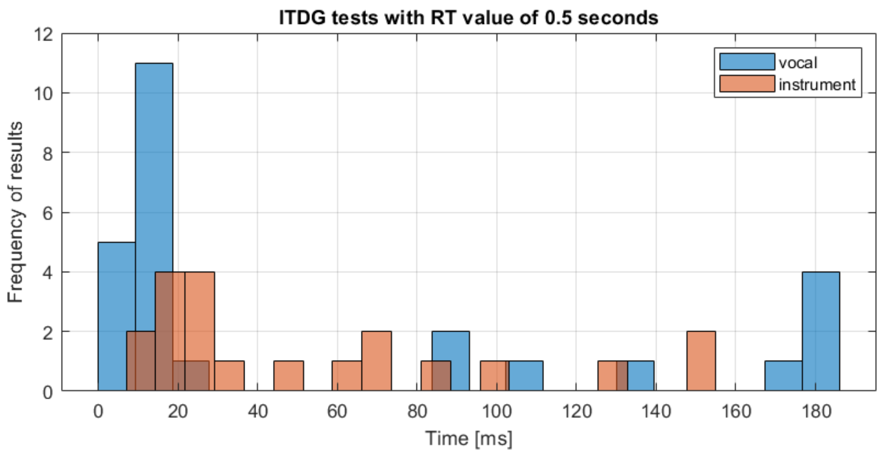


Figure A.1. Histogram of results on the two ITDG tests with RT value of 0.5 seconds.

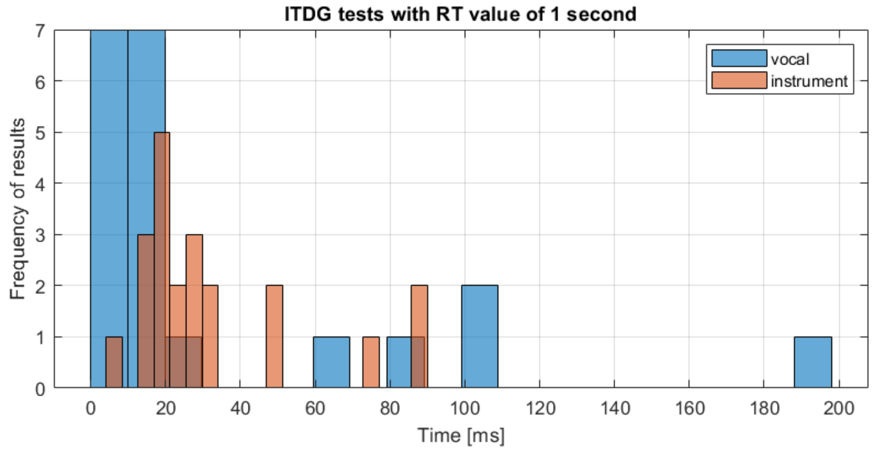


Figure A.2. Histogram of results on the two ITDG tests with RT value of 1 second.

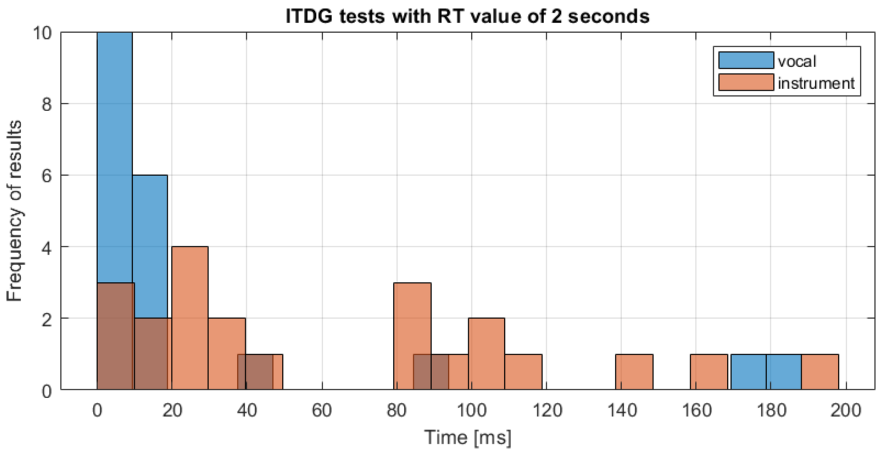


Figure A.3. Histogram of results on the two ITDG tests with RT value of 2 seconds.

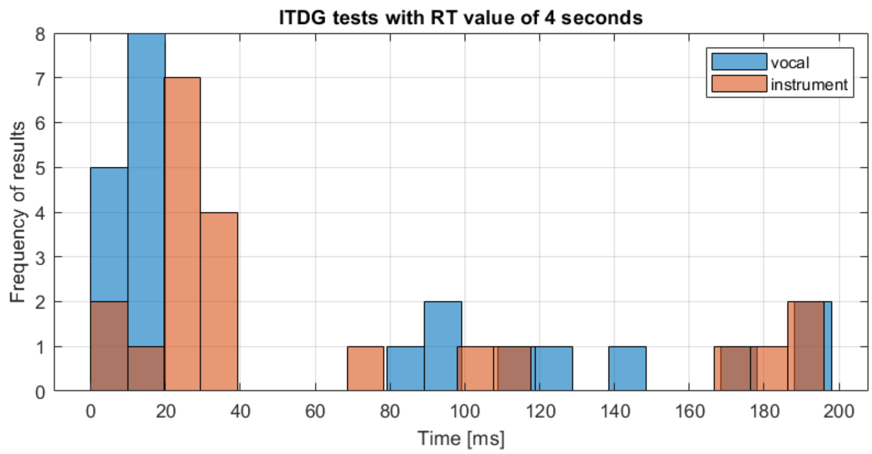


Figure A.4. Histogram of results on the two ITDG tests with RT value of 4 seconds.

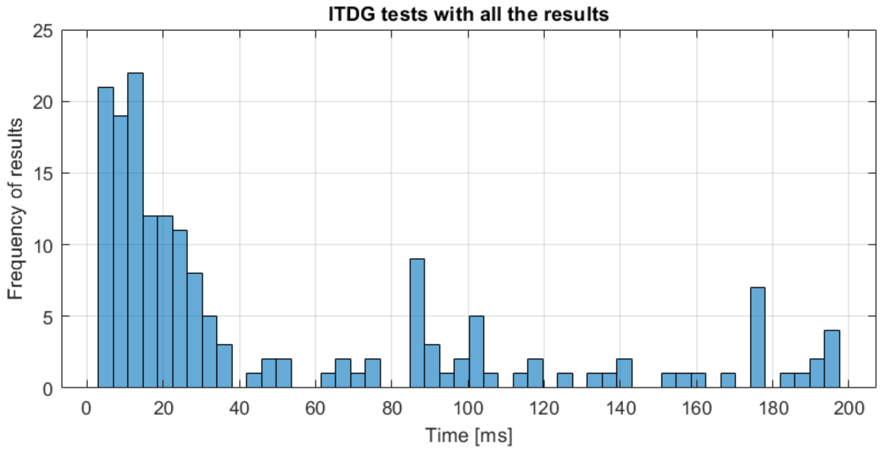


Figure A.5. Results histogram containing the results of all ITDG tests.

RT

In this section there are 4 histograms for the different reverbs used for the RT test and in each one there are the two types of samples used differentiated by colour.

Contrary to the ITDG in this case, a histogram with the sum of all results obtained in all 8 tests has not been added because the perception of RT is dependent on the duration of the reverberation.

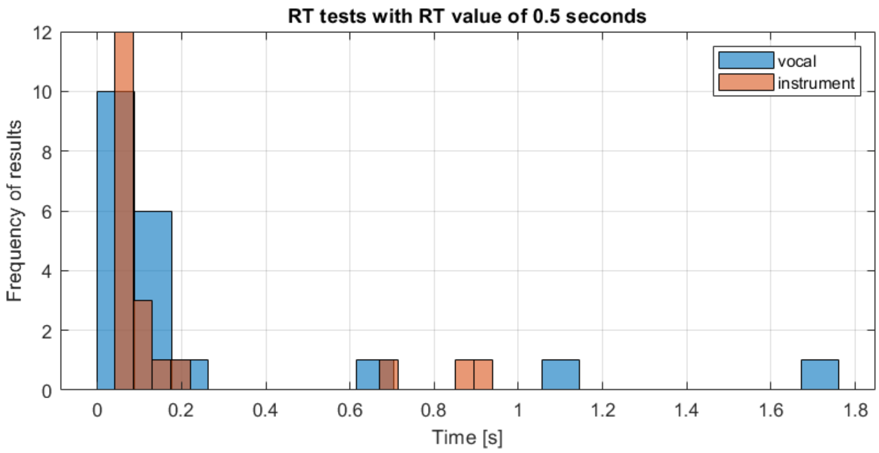


Figure A.6. Histogram of results on the two RT tests with RT value of 0.5 seconds.

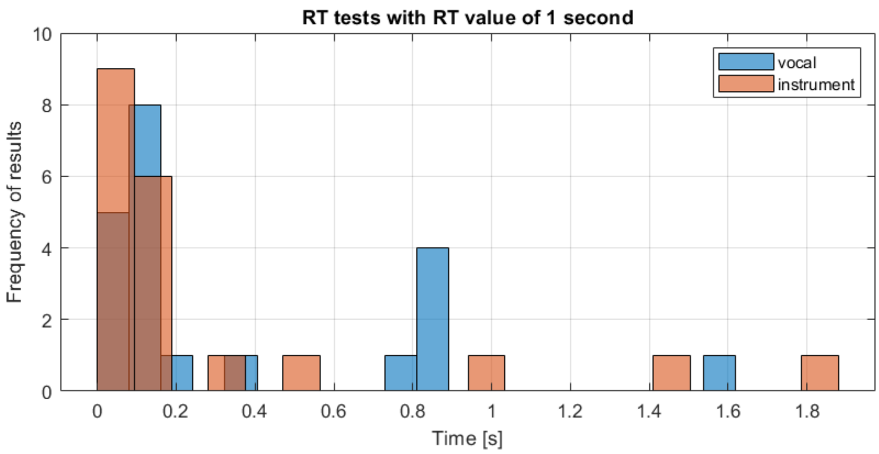


Figure A.7. Histogram of results on the two RT tests with RT value of 1 second.

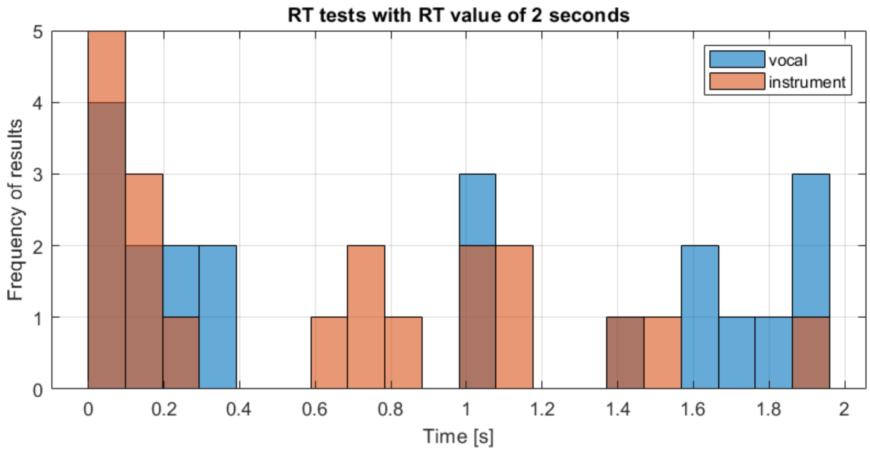


Figure A.8. Histogram of results on the two RT tests with RT value of 2 seconds.

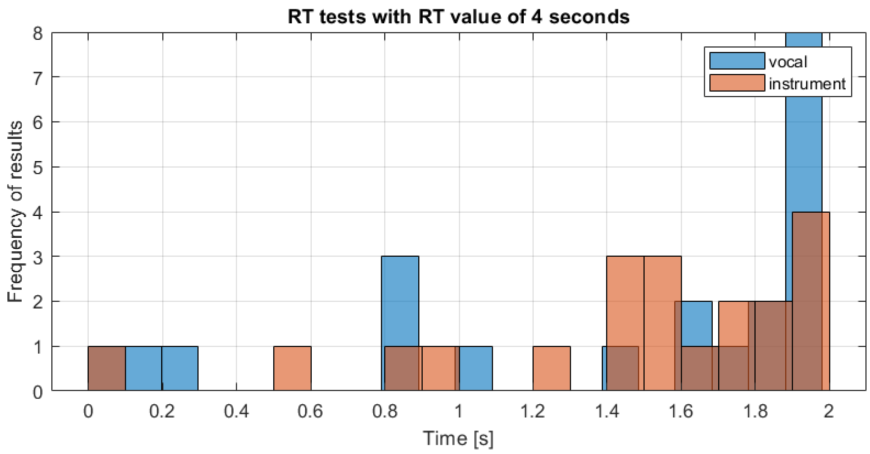


Figure A.9. Histogram of results on the two RT tests with RT value of 4 seconds.

Density

In this section there are two histograms, the first plots the result of the two tests on top of each other to see the differences, the second plots the sum of the two to get an overview of the result.

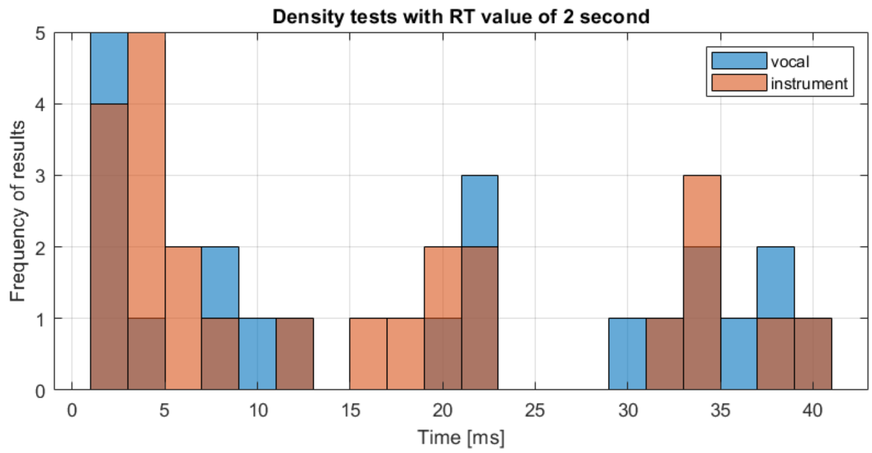


Figure A.10. Histogram of results on the two Density tests with RT value of 2 seconds.

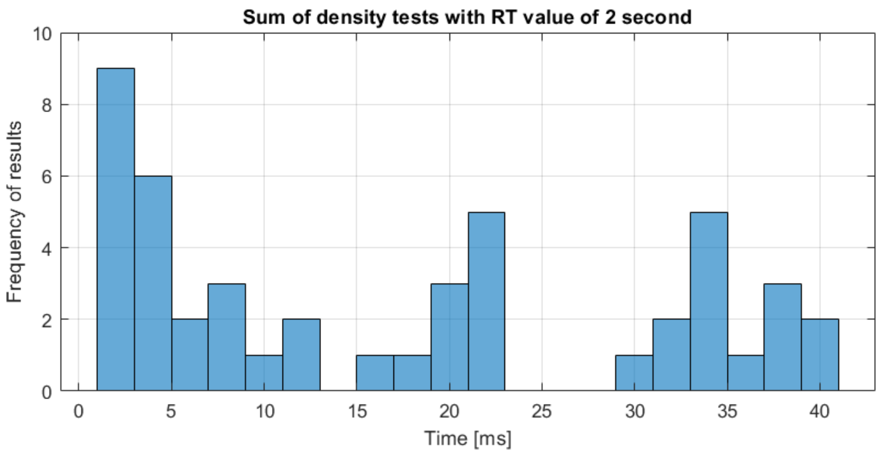


Figure A.11. Results histogram containing the results of all Density tests.

Appendix B

Statistical Analysis

Markov Chain Monte Carlo method

Markov chain Monte Carlo method for ITDG tests

MCMC Summary

		Posterior		95% Credible Interval		Rhat		
Parameter	Mean	Median	SD	Lower	Upper	Point est.	Upper CI	Effecive Sample Size
beta1	4.636	4.613	2.998	-1.126	10.609	1.000	1.000	15000.000
beta2	-1.340	-1.390	6.281	-13.705	10.978	1.000	1.000	14624.146
intercept	44.054	43.918	11.527	21.366	66.727	1.000	1.001	15000.000

Note. The multivariate potential scale reduction factor is estimated at 1.001.

Figure B.1. Summary of ITDG test results obtained from statistical analysis using Markov chain Monte Carlo method.

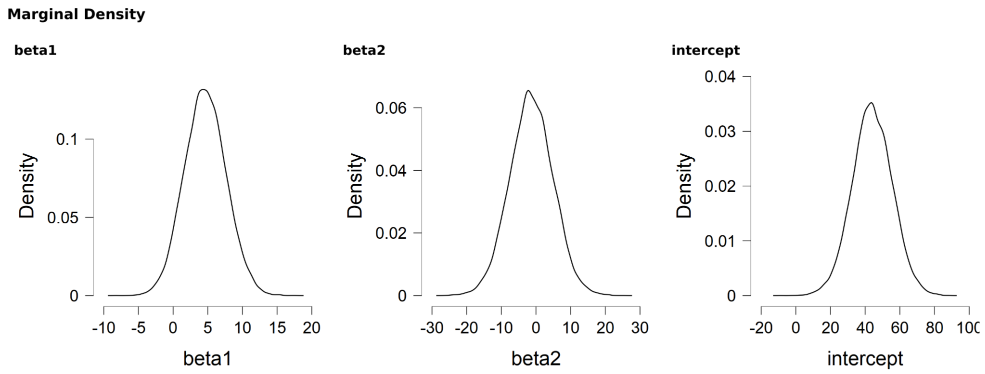


Figure B.2. Plot of the Marginal Density of ITDG test results obtained from statistical analysis using Markov chain Monte Carlo method.

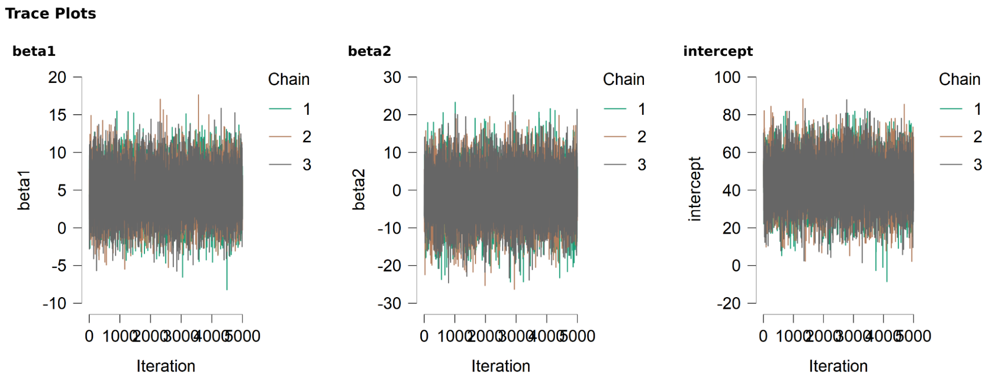


Figure B.3. Trace Plot of ITDG test results obtained from statistical analysis using Markov chain Monte Carlo method.

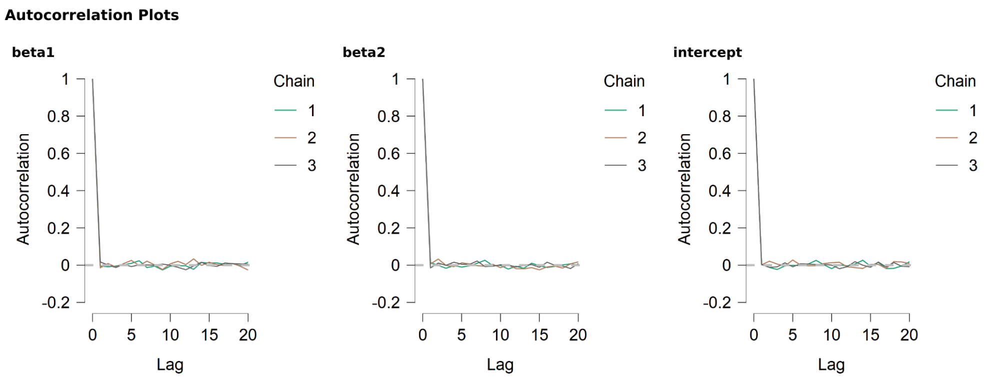


Figure B.4. Autocorrelation Plot of ITDG test results obtained from statistical analysis using Markov chain Monte Carlo method.

Markov chain Monte Carlo method for RT tests

MCMC Summary

		Posterior		95% Credible Interval		Rhat		
Parameter	Mean	Median	SD	Lower	Upper	Point est.	Upper CI	Effecive Sample Size
beta1	0.355	0.355	0.031	0.295	0.417	1.000	1.000	6370.187
beta2	0.063	0.062	0.085	-0.103	0.231	1.002	1.005	6000.000
intercept	-0.066	-0.065	0.146	-0.354	0.222	1.001	1.003	6000.000

Note. The multivariate potential scale reduction factor is estimated at 1.001.

Figure B.5. Summary of RT test results obtained from statistical analysis using Markov chain Monte Carlo method.

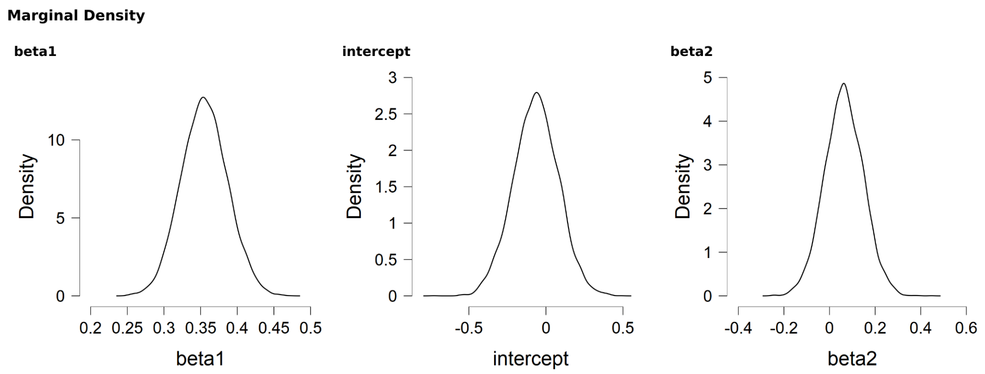


Figure B.6. Plot of the Marginal Density of RT test results obtained from statistical analysis using Markov chain Monte Carlo method.

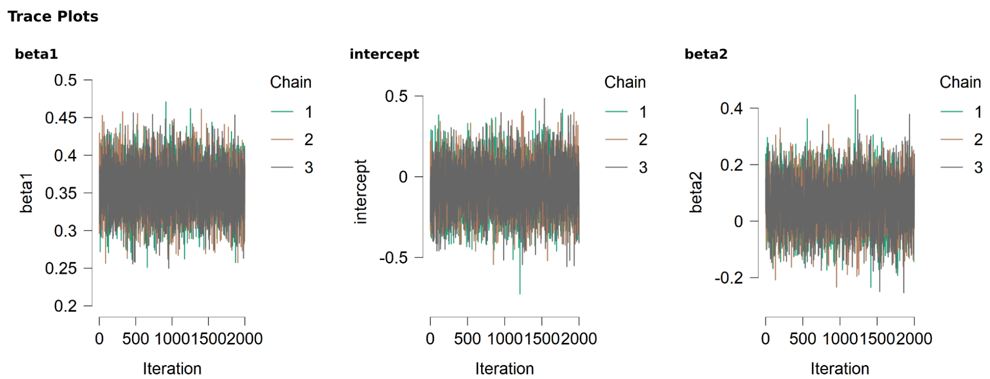


Figure B.7. Trace Plot of RT test results obtained from statistical analysis using Markov chain Monte Carlo method.

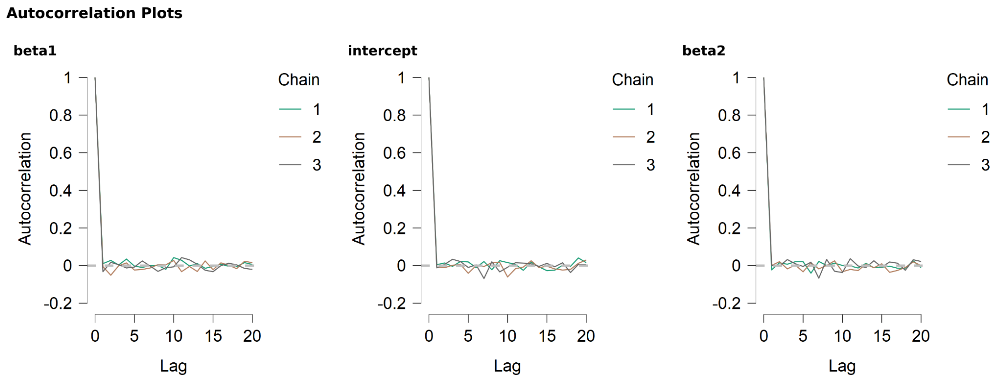


Figure B.8. Autocorrelation Plot of RT test results obtained from statistical analysis using Markov chain Monte Carlo method.

Markov chain Monte Carlo method for Density tests

MCMC Summary

Parameter	Mean	Posterior Median	SD	95% Credible Interval Lower	95% Credible Interval Upper	Rhat Point est.	Rhat Upper CI	Effecive Sample Size
beta2	3.255	3.257	3.699	-4.000	10.467	1.000	1.001	14721.648
intercept	12.411	12.417	5.792	0.916	23.785	1.000	1.001	14513.307

Note. The multivariate potential scale reduction factor is estimated at 1.000.

Figure B.9. Summary of Density test results obtained from statistical analysis using Markov chain Monte Carlo method.

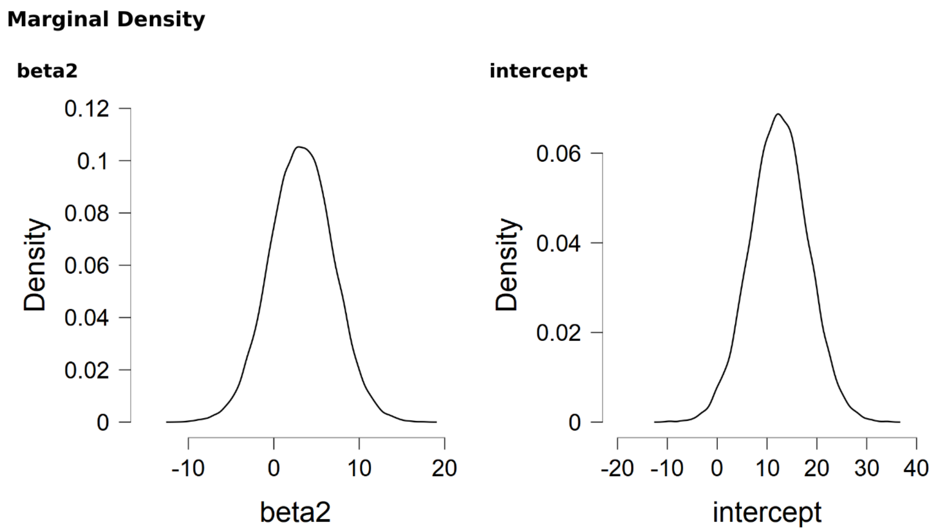


Figure B.10. Plot of the Marginal Density of Density test results obtained from statistical analysis using Markov chain Monte Carlo method.

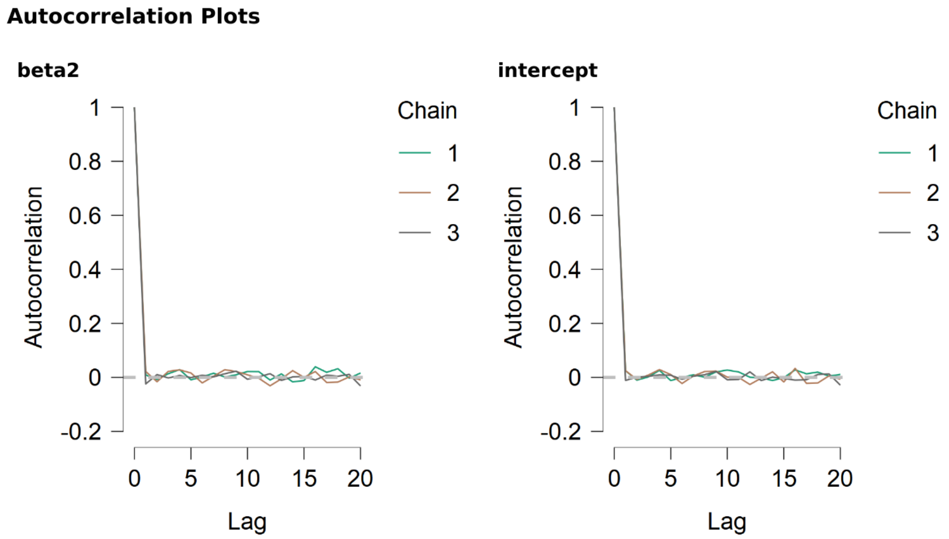


Figure B.11. Autocorrelation Plot of Density test results obtained from statistical analysis using Markov chain Monte Carlo method.

Bayesian Linear Regression method

Bayesian Linear Regression method for ITDG tests

Bayesian Linear Regression

Model Comparison - result

Models	P(M)	P(M|data)	BF_M	BF_{01}	R^2
Null model	0.500	0.722	2.591	1.000	0.000
reverb length	0.500	0.278	0.386	2.591	0.011

Figure B.12. Model comparison table of ITDG test results obtained from statistical analysis using Bayesian Linear Regression method.

Posterior Summary

Posterior Summaries of Coefficients

								95% Credible Interval	
Coefficient	P(incl)	P(excl)	P(incl|data)	P(excl|data)	$BF_{inclusion}$	Mean	SD	Lower	Upper
Intercept	1.000	0.000	1.000	0.000	1.000	51.408	4.349	43.021	59.734
reverb length	0.500	0.500	0.278	0.722	0.386	1.127	2.442	-0.134	7.698

Figure B.13. Posterior summary table of ITDG test results obtained from statistical analysis using Bayesian Linear Regression method.

Bayesian Linear Regression method for RT tests

Bayesian Linear Regression

Model Comparison - result

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Null model	0.500	6.843e -20	6.843e -20	1.000	0.000
reverb.length	0.500	1.000	∞	1.461e +19	0.446

Figure B.14. Model comparison table of RT test results obtained from statistical analysis using Bayesian Linear Regression method.

Posterior Summary

Posterior Summaries of Coefficients

								95% Credible Interval	
Coefficient	P(incl)	P(excl)	P(incl|data)	P(excl|data)	$BF_{inclusion}$	Mean	SD	Lower	Upper
Intercept	1.000	0.000	1.000	0.000	1.000	0.699	0.042	0.617	0.782
reverb.length	0.500	0.500	1.000	0.000	1.461e +19	0.350	0.031	0.289	0.411

Figure B.15. Posterior summary table of RT test results obtained from statistical analysis using Bayesian Linear Regression method.

Appendix C

Comparison of results

ITDG and RT results filtered results

ITDG tests

Initial Time Delay Gap				
type of sample	RT [s]	mean total [ms] + SD	mean without hearing problems [ms] + SD	mean without acoustic knowledge [ms] + SD
vocal	0.5	57.7 + 68.5	57.0 + 69.4	28.4 + 55.2
instrumental		55.4 + 47.2	61.0 + 49.2	49.5 + 36.4
vocal	1	36.1 + 50.3	34.7 + 53.9	20.4 + 31.0
instrumental		32.9 + 23.8	34.3 + 25.5	22.5 + 11.2
vocal	2	32.1 + 54.4	27.9 + 48.8	8.5 + 3.3
instrumental		66.0 + 55.3	64.0 + 58.7	45.7 + 41.6
vocal	4	61.8 + 68.4	50.3 + 59.4	48.6 + 61.6
instrumental		65.2 + 66.2	58.6 + 64.1	37.4 + 29.9
vocal and instrumental	all	51.4 + 57.1	48.5 + 53.6	32.6 + 33.8

Table C.1. Table that contain all the results of ITDG tests with the filtered one.

RT tests

Reverberation Time				
type of sample	RT [s]	mean total [ms] + SD	mean without hearing problems [ms] + SD	mean without acoustic knowledge [ms] + SD
vocal instrumental	0.5	0.256 + 0.434 0.195 + 0.281	0.131 + 0.159 0.180 + 0.285	0.090 + 0.056 0.132 + 0.196
vocal instrumental	1	0.360 + 0.437 0.325 + 0.510	0.330 + 0.367 0.184 + 0.264	0.192 + 0.257 0.178 + 0.299
vocal instrumental	2	0.885 + 0.752 0.662 + 0.582	0.654 + 0.689 0.568 + 0.529	0.799 + 0.681 0.414 + 0.434
vocal instrumental	4	1.411 + 0.655 1.473 + 0.507	1.368 + 0.695 1.480 + 0.501	1.453 + 0.600 1.599 + 0.300

type of sample	RT [s]	mean total [%]	mean without hearing problems [%]	mean without acoustic knowledge [%]
vocal and instrumental	all	38.4%	30.8%	27.3%

Table C.2. Table that contain all the results of RT tests with the filtered one.

Bibliography

[1] Y Ando and AH Marshall. *Concert Hall Acoustics by Y. Ando*. 1986.

[2] Octávio Inácio. "FUNDAMENTALS OF ROOM ACOUSTICS". In: *Lda. R. Dr. Carlos Pires Felgueiras* (Aug. 2), pp. 173–6. DOI: 10.1007/978-3-540-92723-5_7.

[3] M Barron. "Growth and decay of sound intensity in rooms according to some formulae of geometric acoustics theory". In: *Journal of Sound and Vibration* 27.2 (1973), pp. 183–196.

[4] Heinrich Kuttruff. *Room acoustics*. Crc Press, 2016.

[5] Werner Helmut Haas. *Taschen-Lexikon der Abkürzungen*. WH Haas, 1949.

[6] Lothar Cremer. "Law of the first wave front". In: *Journal of the Audio Engineering Society* 25.6 (1977), pp. 420–422.

[7] Y. Ando and A. H. Marshall. "Concert Hall Acoustics by Y. Ando". In: *The Journal of the Acoustical Society of America* 79.2 (1986), pp. 582–583. DOI: 10.1121/1.393515. eprint: https://doi.org/10.1121/1.393515. URL: https://doi.org/10.1121/1.393515.

[8] Leo Beranek and Daniel W Martin. *Concert & opera halls: how they sound*. 1996.

[9] A Cocchi, A Farina, and L Rocco. "Reliability of scale-model researches: a concert hall case". In: *Applied Acoustics* 30.1 (1990), pp. 1–13.

[10] S SATO. "On the Subjective Evaluation of Source Locations on the stage by Listeners, Introducing Interaural Time Delay of IACC". In: *Proc. Intern. Sympo. Music & Concert Hall Acoust., Kirishima* (1995).

[11] Yoichi Ando et al. "Acoustic design of a concert hall applying the theory of subjective preference, and the acoustic measurement after construction". In: *Acta Acustica united with Acustica* 83.4 (1997), pp. 635–643.

[12] Yoichi Ando. "Subjective preference in relation to objective parameters of music sound fields with a single echo". In: *The Journal of the Acoustical Society of America* 62.6 (1977), pp. 1436–1441.

[13] Yoichi Ando, M Okura, and K Yuasa. "On the preferred reverberation time in auditoriums". In: *Acta Acustica united with Acustica* 50.2 (1982), pp. 134–141.

[14] Yoichi Ando, Toshiyuki Okano, and Yoshitaka Takezoe. "The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields". In: *The Journal of the Acoustical Society of America* 86.2 (1989), pp. 644–649.

[15] Leo L Beranek. “Subjective rank-orderings and acoustical measurements for fifty-eight concert halls”. In: *Acta Acustica united with Acustica* 89.3 (2003), pp. 494–508.

[16] Leo L Beranek. “Music, acoustics, and architecture”. In: *Bulletin of the American Academy of Arts and Sciences* 45.8 (1992), pp. 25–46.

[17] Helmut V Fuchs. *Applied acoustics: Concepts, absorbers, and silencers for acoustical comfort and noise control: Alternative solutions-Innovative tools-Practical examples*. Springer Science & Business Media, 2013.

[18] Francesco Martellotta. “The just noticeable difference of center time and clarity index in large reverberant spaces”. In: *The Journal of the Acoustical Society of America* 128 (Aug. 2010), pp. 654–63. DOI: 10.1121/1.3455837.

[19] Helmut A Müller. *Principles and applications of room acoustics*. Applied Science, 1982.

[20] Theodoros I Niaounakis and William J Davies. “Perception of reverberation time in small listening rooms”. In: *Journal of the Audio Engineering Society* 50.5 (2002), pp. 343–350.

[21] Matthew G Blevins et al. *Quantifying the just noticeable difference of reverberation time with band-limited noise centered around 1000 Hz using a transformed up-down adaptive method*. 2013.

[22] CCJM Hak, RHC Wenmaekers, and LCJ Van Luxemburg. “Measuring room impulse responses: Impact of the decay range on derived room acoustic parameters”. In: *Acta Acustica united with Acustica* 98.6 (2012), pp. 907–915.

[23] Acoustics ISO. “Measurement of room acoustic parameterspart 1”. In: *ISO Std* (2009).

[24] Hans-Peter Seraphim. “Untersuchungen über die Unterschiedsschwelle exponentiellen Abklingens von Rauschbandimpulsen”. In: *Acta Acustica united with Acustica* 8.4 (1958), pp. 280–284.

[25] J Atagi, R Weber, and V Mellert. “Effect of modulated delay time of reflection on autocorrelation function and perception of echo”. In: *Journal of sound and vibration* 258.3 (2002), pp. 443–450.

[26] Toshiyuki Okano. “Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections”. In: *The Journal of the Acoustical Society of America* 111 (Feb. 2002), pp. 217–29. DOI: 10.1121/1.1426374.

[27] RJ Hawkes and H Douglas. “Subjective acoustic experience in concert auditoria”. In: *Acta Acustica united with Acustica* 24.5 (1971), pp. 235–250.

[28] Marshall Long. *Architectural acoustics*. Elsevier, 2005.

[29] Heiko Gölzer and Michael Kleinschmidt. “Importance of early and late reflections for automatic speech recognition in reverberant environments”. In: *Proc. Elektronische Sprachsignalverarbeitung (ESSV)* (2003), pp. 98–105.

[30] EN ISO and BRITISH STANDARD. *Acoustics–Measurement of room acoustic parameters*. 2008.

[31] HCCH Levitt. “Transformed up-down methods in psychoacoustics”. In: *The Journal of the Acoustical society of America* 49.2B (1971), pp. 467–477.

[32] Eckart Scheerer. “The unknown Fechner”. In: *Psychological research* 49.4 (1987), pp. 197–202.

[33] B.M. Gibbs and D. Jones. “A Simple Image Method for Calculating the Distribution of Sound Pressure Levels within an Enclosure”. In: *Acta Acustica united with Acustica* 26 (Jan. 1972).

[34] Jont Allen and David Berkley. “Image method for efficiently simulating small-room acoustics”. In: *The Journal of the Acoustical Society of America* 65 (Apr. 1979), pp. 943–950. DOI: 10.1121/1.382599.

[35] Stephen G McGovern. “Fast image method for impulse response calculations of box-shaped rooms”. In: *Applied Acoustics* 70.1 (2009), pp. 182–189.

[36] Marco Faria, Frederico Oliveira, and Gilberto Pimentel-Junior. *ACOUSTIC EMISSION TESTS ON THE ANALYSIS OF CRACKED SHAFTS OF DIFFERENT CRACK DEPTHS*. Dec. 2015. DOI: 10.20906/CPS/COB-2015-1434.