POLITECNICO DI TORINO

Master's Degree in Cinema & Media Engineering

Master's Degree Thesis

Real-Time Auralization of Large Spaces and its Effects on Singing Performance

Supervisors
Prof. Arianna ASTOLFI
Prof. Louena SHTREPI
Prof. Alessio CARULLO
Dr. Angela GUASTAMACCHIA

Candidate
Nicolò SOMÀ

April 2025

Abstract

Many studies have examined how singers adapt their vocal delivery to the acoustic properties of different performance spaces. However, most virtual acoustic simulations rely on binaural rendering rather than multi-speaker systems. This thesis explores whether a similar adaptive behavior occurs when reverberant environments are reproduced through real-time auralization. A real-time convolution system was developed within the Audio Space Lab (ASL) at the Department of Energy, Politecnico di Torino, using 3rd Order Ambisonics impulse responses from four contemporary churches in Milan, Italy. The laboratory features a 16-loudspeaker spherical array, allowing singers to experience immersive acoustic simulations.

Six professional singers originally performed in these churches, where the acoustic properties were measured following ISO 3382-1, and their vocal delivery was analyzed. Subjective feedback on the ease of singing was also collected. A questionnaire has also been used to collect feedback on vocal comfort while performing in each on those environments.

In the ASL, amateur singers were invited to perform freely in the auralized environments and rate their preferences through a subjective questionnaire similar to the one provided in the actual churches. Also, their vocal performances were recorded while singing a well-known, simple melody, and their delivery parameters were extracted and compared to those of the professional singers. The subjective ratings of the amateur singers in the virtual environments were also compared to the impressions of the professionals who had performed in the actual churches. Finally, vocal parameters were calculated and compared to the ones from the professional singers, along with the preference charts obtained from both groups.

This study aimed to evaluate the feasibility of real-time convolution within the ASL and investigate how amateur singers perceive and adapt to auralized spaces. The findings provide insights into the effectiveness of multi-speaker auralization in vocal performance research.

A Mamma e Papà, che mi hanno permesso di arrivare fino a qui.
A Gabriele e Carlotta, che sono la mia forza.
A Nonno Roberto, Nonno Pierangelo, Nonna Ernestina e Nonna Maria,
che mi hanno insegnato a non mollare mai.

"La paura, ogni paura, è una perdita di tempo. Come i rimpianti."
Indíra Gándhí

Table of Contents

List of Tables

List of Figures

Acronyms

3OA	Third Order Ambisonic
ASL	Audio Space Lab
C_{80}	Clarity
CPPS	Cepstral Peak Prominence Smoothed
D_{50}	Definition
$\mathrm{DT}_{40,ME}$	Decay Time at the ears
EDT	Early Decay Time
BR	Bass Ratio
FCP	Formant Cluster Prominence
F_0	Fundamental Frequency
G_{RG}	Room Gain
HRTF	Head-Related Transfer Function
IACC	Inter-Aural Cross-correlation Coefficient
IACC_{early}	Early Inter-Aural Cross Correlation
IACC_{late}	Late Inter-Aural Cross Correlation
IID	Inter-aural Intensity Difference
IR	Impulse Response
ITD	Inter-aural Time Difference

LTAS	Long-Term Average Spectrum
OBRIR	Oral-Binaural Room Impulse Response
RT	Reverberation Time
SMA	Spherical Microphone Array
SPL	Sound Pressure Level
SPR	Singing Power Ratio
ST_V	Vocal Support
T_{20}	Reverberation Time - 20
T_{30}	Reverberation Time - 30
TD	Time Dose
T_s	Center Time
TR	Treble Ratio

Chapter 1

Introduction

The concept of auralization has been explored for nearly a century. From early experimental setups—where scaled room models were used to simulate the acoustics of real spaces—to the first implementations utilizing magnetic tape recorders, the fundamental goal of auralization has remained clear in the minds of researchers: to model a sound field in which an arbitrary sound signal is processed and subsequently reproduced through an appropriate audio system.

However, it was only with the advent of modern computing technology that realistic auralization became feasible. Both room acoustics simulators and convolution processors progressively evolved, culminating in the formal introduction of the term "auralization" in 1992 [1]. Michael Vorländer, one of the leading scholars in the field of virtual acoustics, described auralization as the auditory counterpart of visualization:

> *In acoustics, auralization occurs when acoustic effects, primary sound signals or means of sound reinforcement or sound transmission, are processed into an audible result.* [2]

As stated in a more recent paper [3], the auralization process consists in three steps: the recording of the audio source material, which would be then convolved with the response of the room of interest, measured in a given position; the result should be rendered via proper sound rendering systems.

A crucial and computationally intensive step in the auralization process is convolution, which enables the spatial placement of sounds within measured acoustic environments. Convolution is known for its high computational cost, requiring optimized algorithms to ensure a balance between output quality and minimal processing latency.

The advancement of computational power was pivotal in overcoming one of the major limitations of early auralization systems: latency. Until then, auralization had been primarily used for simulating the placement of sound sources within

virtual rooms, but real-time interaction between a performer and the recreated space remained unachievable.
Another fundamental aspect is the accurate reproduction of acoustic feedback to a performer. Both the latency and the acoustic properties of the system must be precisely calibrated to achieve an immersive experience in the virtual acoustic environment.
The objective of this thesis was to design and implement a real-time auralization system capable of recreating the acoustic behavior of reverberant spaces using the 3rd Order Ambisonic (3OA) loudspeaker array in the Audio Space Lab at Politecnico di Torino. Specifically, the focus was on evaluating the feasibility of real-time auralization for singers, allowing them to receive acoustical feedback as if they were performing in a physical space.
The target environments selected for auralization were four modern churches in Milan, Italy, chosen for their distinctive architectural and material characteristics, which resulted in four unique acoustic profiles. The 3OA Impulse Responses (IR) of these spaces were measured and integrated into the convolution engine of the auralization system. Subsequently, a subjective experimental procedure was conducted with amateur singers to assess the perceived realism of the generated acoustic feedback.
Furthermore, particular attention was dedicated to analyzing whether the vocal parameters of the singers were influenced by the acoustic properties of the auralized spaces. While previous studies have attempted to correlate room acoustics with vocal adaptation, findings remain scarce and, in some cases, contradictory. To address this gap, recordings of a soloist performing in the actual churches were analyzed, and the same vocal features were extracted from recordings of amateur singers experiencing the auralized environments. The singers' behavior in both scenarios was then correlated with the acoustic characteristics of the real churches to determine whether consistent patterns and adaptations could be observed.

1.1 Auralization for singing: state of the art

Auralization is applied in a wide variety of simulation contexts; however, few applications have been specifically designed for the field of singing research.
Yadav et al. [4] developed a real-time auralization system using Max/MSP software. Oral-Binaural Room Impulse Responses (OBRIR) were recorded by emitting a sweep stimulus from the mouth simulator of an artificial head and capturing the resulting sound at both ear positions using two microphones. During each recording, the artificial head was systematically rotated to generate a comprehensive set of OBRIRs for both ears. This approach allowed the authors to design a simulative

auralization setup within an anechoic chamber, enabling individuals to perceive their own voice through a pair of ear-loudspeakers. A head-tracking sensor was then employed to dynamically select the appropriate pair of OBRIRs for real-time convolution. Additionally, the authors reported that the use of ear-loudspeakers had a negligible impact on the sound level reaching the ears.
Miranda Jofre et al. [5] developed a real-time auralization system to investigate the influence of temporal and spatial characteristics of stage acoustics on the performance of a trained singer. Early reflections were synthesized by adjusting their spatial and temporal properties to match the desired acoustic conditions, while the late reverberant tail—recorded using an artificial head simulator—was subsequently concatenated. The results showed no clear pattern in the singer's preferences regarding variations in these acoustic components.

Significant advancements have been made in the field of auralization in recent years, with particular emphasis on Geometrical Acoustic (GA) simulations. This approach addresses the challenge of recording IRs in real spaces by relying on the accuracy of simulated models. However, such models must be meticulously designed and calibrated to ensure their reliability in replicating real acoustic environments, especially when compared to the direct recording and reproduction of actual room responses.
Postma et al. [6, 7, 8] proposed a methodology for creating and calibrating GA model-based auralizations, particularly for the case of churches. The calibration was tested using objective acoustic parameters, with the assumption that realism is achieved if reverberation and clarity fall within 1 JND of the measured values. Subjective listening tests were then conducted, using stimuli that were properly convoluted and reproduced through headphones. However, the authors themselves acknowledged that the methodology used in this preliminary study may not be applicable to environments with well-distributed absorption and scattering-independent reverberation times.
Another calibration procedure is proposed in [9]. In this paper, *Mullins et al.* introduced a methodology for calibrating the output of the auralization system based on the Stage Support (ST_{early}) provided in both actual and virtual environments. The formula to compute ST_{early} was slightly modified to exclude direct sound from the calculation, allowing the authors to obtain a measure of late feedback energy. This measure can be used to match the levels between real and virtual environments. The impulse responses needed for these calculations were obtained using a dummy head simulator, as the auralization output was rendered through head-mounted loudspeakers.

These workflows work well when the acoustical parameters of real environments are measurable and comparable to those obtained after auralization. However, this could be a significant limitation in cases where the environment in question is no longer measurable.

Katz et al. focused their research on the context of auralization applied to archeo-acoustics. This fascinating field pays particular attention to the study of the acoustical characteristics of ancient rooms and buildings that have been damaged or destroyed over time. Specifically, the main focus of their "*EVAA*" project [10] was to understand how the acoustics of historical venues influenced the development of musical instruments, traditional music composition techniques, and the performance of musicians across different contexts.

Many eperiments were conducted in the field of GA applied for auralization in churches, using various tecniques and experimental setups.

Thery et al. [11] conducted a perceptual investigation to assess the differences in the reproduction of a set of auralized stimuli rendered via a loudspeaker array and binaural reproduction via headphones. For the auralization process, two similarly sized, small rooms were chosen for their acoustical peculiarities and differences. These rooms were modeled using GA software. Various anechoic stimuli were selected based on their timbral and stylistic characteristics. The scope of their study was to assess the reliability of using auralization as a tool for decision-making during the room design process. Ideally, the two reproduction methods should not affect the characteristics of the result, enabling the user to make design decisions independently of the reproduction device in use. The results showed a slight difference in the ratings of *Apparent Source Width* (*ASW*) and *Listener Envelopment* (*LEV*), which were attributed to less-than-ideal anechoic conditions in which the loudspeaker array was installed. Consequently, the authors recommended using headphones over loudspeakers if ideal conditions are not achievable, especially during detailed listening.

Eley et al. [12] auralized the performance of a four-member ensemble using both a 32-channel loudspeaker array and binaural reproduction via open-back headphones. The performers were free to move while wearing a head-mounted motion capture sensor, allowing their directivity to be adapted during the simulation. The singers reported a general preference for the loudspeaker array auralization system, as reproduction through headphones, although perceived as more realistic, negatively impacted their ability to hear themselves while singing. The aim of this study was to realistically reproduce the acoustical features of the Cathedral of Notre-Dame de Paris (Paris, France) prior to the 2019 fire. For the experimental tests, only singers with a solid familiarity with the acoustics of the church were chosen. Their feedback on the plausibility of the auralized environment and its similarity to the actual acoustics of the original church was collected, with an overall positive response.

DeMuynke et al. [13] conducted a similar experiment, where a GA model of the Great Chapel of the Palais des Papes (Avignon, France) was created and calibrated. The objective of the study was to investigate the influence of the cathedral's acoustics on the performance of a four-member ensemble. The auralization system was designed to perform binaural, real-time auralization of all the members simultaneously via open-back headphones. The direct sound from the other singers could reach the ears directly, while the room feedback was simulated through the virtual acoustic environment. Two experimental setups were conducted, using two different church configurations corresponding to the medieval and modern states of the church. The singers' subjective evaluations were gathered via a rating questionnaire, which revealed a consensus regarding the reverberation differences: as expected, the ancient church was much more reverberant than the modern one. The participants provided significant negative feedback regarding the use of the headphones, which they reported altered the direct sound from the other members of the ensemble.

Auralization may be a powerful tool during the design process; however, professionals still find it challenging to implement it in their workflow due to a lack of resources, time, or the necessary skills to produce affordable and representative results [14].

1.2 Voice & Room Acoustics Parameters

Lastly, one of the objectives of the present thesis is to identify a common pattern that may correlate changes or adaptations in voice emission with the acoustics of a venue, as described by its acoustical parameters. While evidence has been found of a negative correlation between speaking comfort and reverberation [15, 16, 17], few studies have investigated the relationship between room acoustics and singing voices.

Investigations have been conducted regarding the behavior of instrumentalists in response to the feedback provided by performance venues. In [18], choir members were observed to increase the intensity of their emission as the reverberation of the venue decreased. In [19], more than 50% of the variance in the execution of a musical piece by a renowned cellist was found to be correlatable with room acoustics.

In [20], a real-time auralization setup with six loudspeakers in an anechoic chamber was used to measure the changes in the performance of four professional instrumentalists and a baritone singer with operatic professional training. The findings underlined a correlation between loudness and vibrato intensity with room response and reverberation, respectively.

In [21, 22], *Luizard et al.* found a sparse correlation between the performance of four operatic singers and the room acoustical features of eight different venues, while singing a well-known piece within a comfortable range. While no clear patterns were found considering the mean of all the singers together, relationships were

observed between individuals and room acoustics features, leading the authors to conclude that each singer developed a subjective way to adapt to the feedback provided by each room. However, generalizing this adaptation across singers was considered beyond the scope of the studies, as precise experimental setups, including a larger number of singers and a specific repertoire, would be required. Additionally, the time passing between recordings in different performance rooms should be considered as a relevant aspect, as psychological and physical changes may occur.
In [23], *Bottalico et al.* focused on vibrato rate, vibrato extent, and pitch accuracy, the latter being evaluated based on notes extracted from the central part of precise sections of the musical score, which was identical for all subjects. Relevant relations were found between voice parameters and room acoustics parameters.

Following these findings, *Redman et al.* [24] proposed three perceptual parameters. Two of them, *Room Supportiveness* and *Room Noiselessness*, were reported to be correlated with singers' likability ratings; *Room Timbre*, linked to BR, was not found to contribute significantly to the singers' preferences, in contrast to findings in [21].

In conclusion, common patterns of adaptation to room acoustics across different venues have yet to be clearly identified, as these adaptations remain largely subjective. Additionally, a standardized set of parameters for describing the vocal features of the singing voice is still lacking. While this thesis did not aim to resolve these gaps, it focused on examining whether similar adaptation mechanisms could be observed between professional and amateur singers.

Chapter 2

Materials And Methods

This study is based on recordings and measurements made in both real and auralized environments. Vocal parameters were extracted and analyzed to identify potential correlations with room acoustics parameters and to determine whether these relationships emerged consistently in both conditions.

In this chapter, the four churches are presented, along with the ensemble Faber Teater. The measurement procedures and the room acoustics parameters of interest are then outlined. Then, the ASL reproduction system is described, along with the omnidirectional and binaural acoustical parameters used to characterize the auralized output signal in this environment. Finally, the voice parameters adopted in the analysis are presented.

2.1 The Four Churches

The performance of the Faber Teater ensemble was recorded during their Stabat Mater tour, which took place from March to April 2022. The four churches considered in this study were therefore selected by the ensemble itself, with no influence from the measurement team in the selection process. The four churches were:

- Church of Santa Gianna Beretta Molla, Trezzano sul Naviglio (MI)
- Church of San Giovanni Battista alla Creta, Milan (MI)
- Church of San Nicolao della Flue, Milan (MI)
- Church of Santi Giovanni Battista and Paolo, Milan (MI)

These churches are all located in the territory of the metropolitan city of Milan, Italy. **Figure 2.1** presents a map with the locations of these sacred buildings.

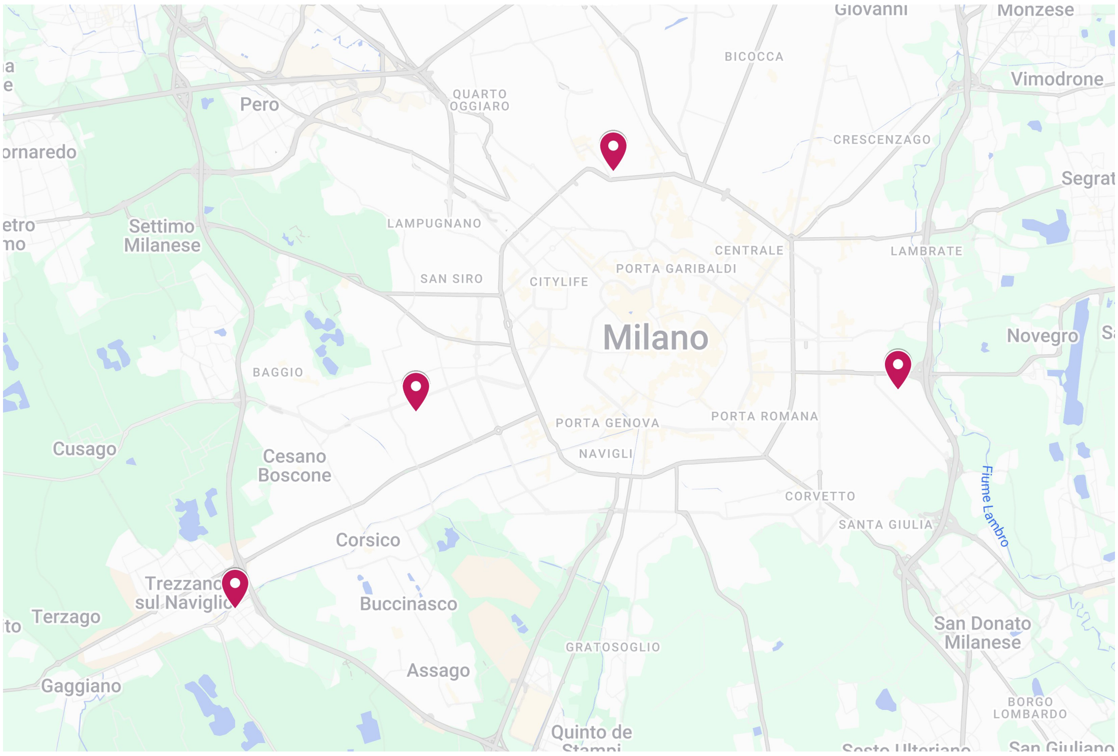
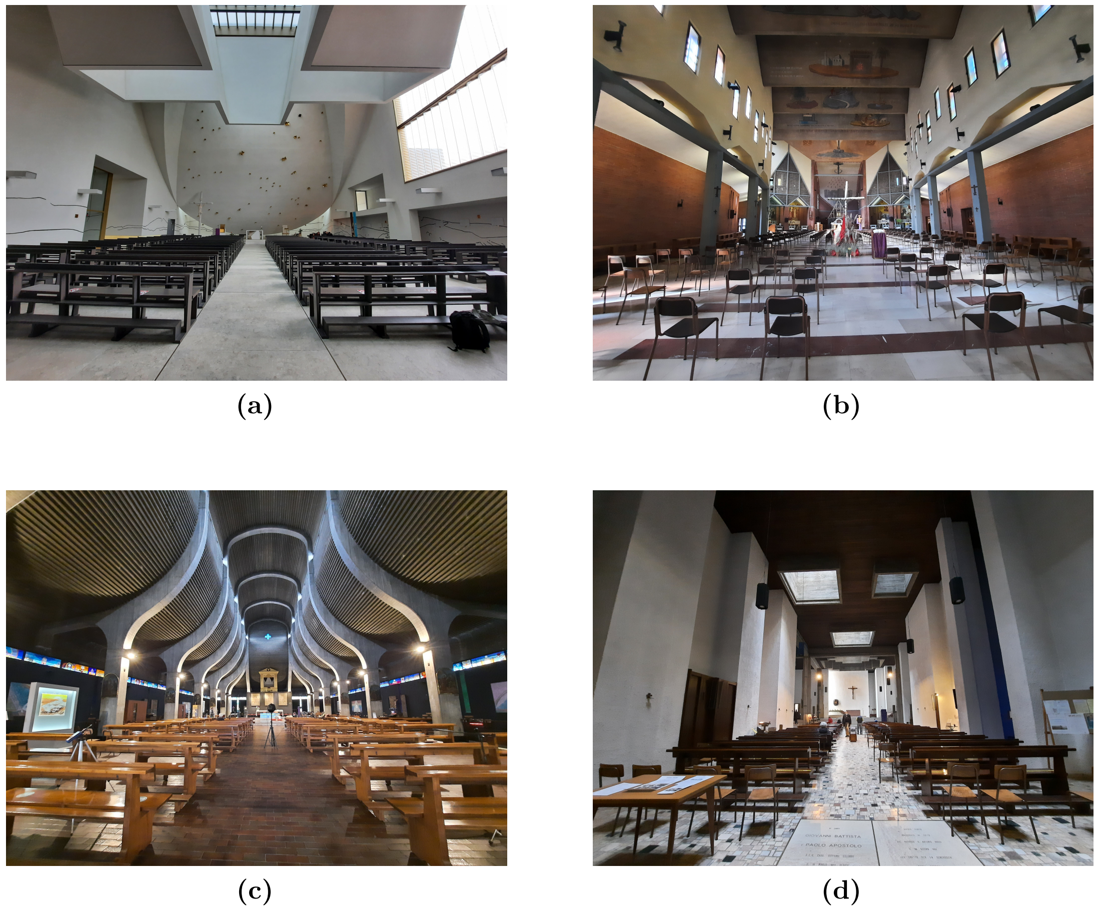


Figure 2.1: Position of the churches

All of the churches were built between the second half of the 20^{th} century and the early 21^{st} century, following a modern architectural style. **Figure 2.2** presents photographs of their interiors, while **Figure 2.3** shows their floor plans.

For the sake of simplicity, the churches will be referred to using numerical labels from this point forward. The numbering follows the order in which the measurements were taken.

The churches exhibit irregular, interconnected volumes accessible from the main nave through portals or openings between columns. Additionally, some of them — particularly Church 1 — have undergone acoustic treatment, with multilayer absorptive materials applied to the walls, which results in peculiar acoustics characteristics. Moreover, Church 3 features a distinctive ceiling over its lateral naves, characterized by a rounded shape with irregularities that may contribute to a scattering effect on reflected sound. A detailed characterization of the acoustic properties of the construction materials and structures was not conducted, as it was considered beyond the scope of this study. However, room acoustics parameters were collected, as described in [25], following the guidelines specific for measuremnts in churches given by [26].



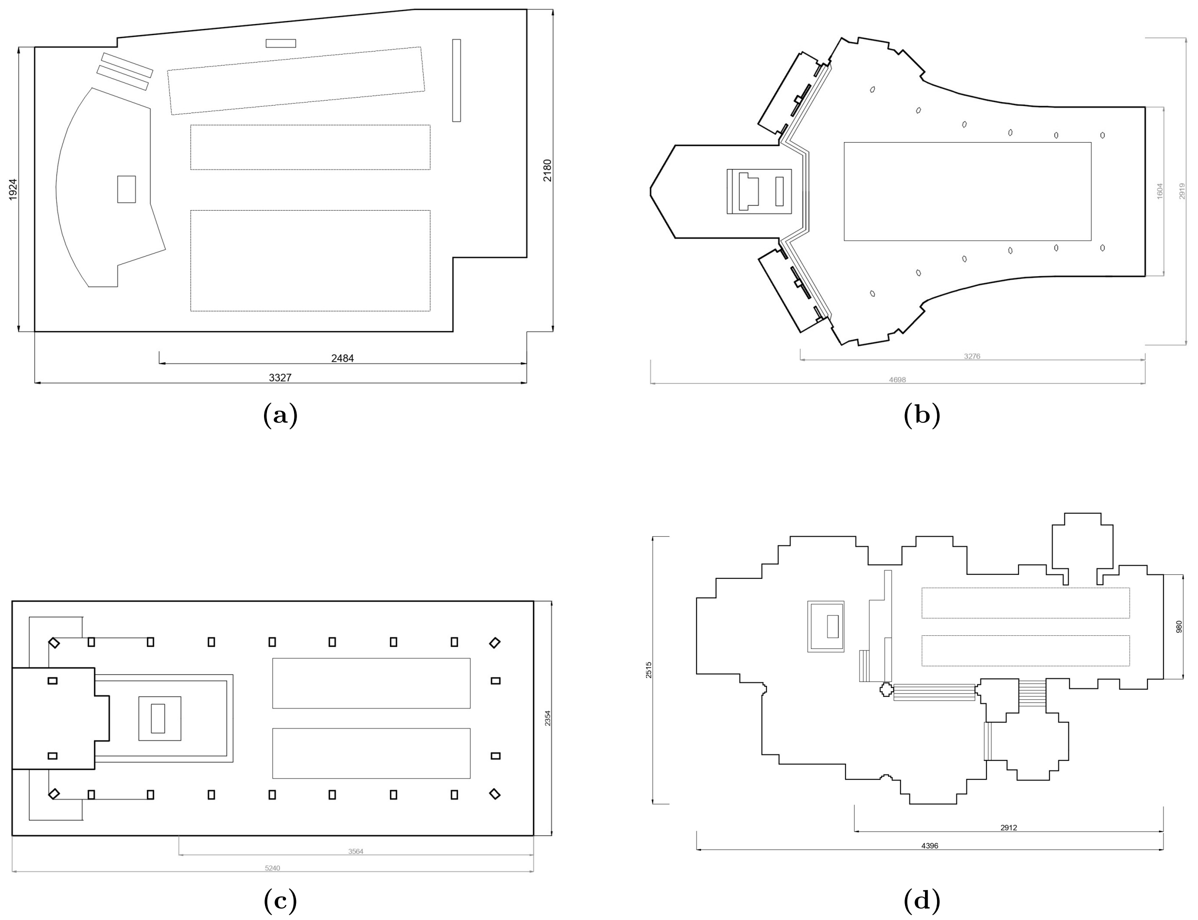
(a)

(b)

(c)

(d)

Figure 2.2: View of the interior of the churches: a) Church 1: Santa Gianna Beretta Molla; b) Church 2: San Giovanni Battista alla Creta; c) Church 3: San Nicolao della Flue; d) Church 4: Santi Giovanni Battista and Paolo.



(a)

(b)

(c)

(d)

Figure 2.3: Plans of the churches: a) Church 1: Santa Gianna Beretta Molla; b) Church 2: San Giovanni Battista alla Creta; c) Church 3: San Nicolao della Flue; d) Church 4: Santi Giovanni Battista and Paolo.

2.2 Faber Teater

Faber Teater [27] is a collective of “theatrical artisans” founded in 1997 and based in the province of Turin, Italy. Originally focused on exploring new ways of experiencing theatrical performances, the group has consistently sought to create an immersive and emotionally engaging connection with the audience. Their approach to theater emphasizes playfulness, self-exploration, and the dynamic relationship between actors and spectators. Each performance becomes an opportunity to share emotions, foster enjoyment, and momentarily escape the complexities of everyday life. In all their works, particular attention is given to the role of the audience and, notably, to the significance of the venue in shaping the theatrical experience.

Always guided by the principles of self-training and pedagogy—where actors independently choose their preferred methods of training—the group, since 2004, has increasingly focused on the vocal dimension of their performances. This shift was influenced by the teachings of Antonella Talamonti, whose guidance allowed them to deepen their understanding of music and vocal pedagogy.

From this collaboration, the creation of *Stabat Mater - Creazione per 6 voci e un duomo* (*Stabat Mater - A Creation for 6 Voices and a Cathedral*) emerged in 2007. This work is a musical-theatrical performance, described as an “emotional and acoustical experience,” in which themes such as loss, death, injustice, the need for consolation, and the act of sharing are explored. The performances took place in several sacred spaces across Northern Italy, where the “singing stone” became an active participant in conveying the performers’ emotions to the audience.

The original compositions, crafted by director Talamonti, are inspired by the oral Italian tradition of chants associated with the Christian Holy Week liturgy, adapted to suit the vocal characteristics of the Faber ensemble. Various languages and dialects were incorporated into the compositions, creating a captivating work that blends theatrical and musical elements in a unique way.

An important part of the preparation process for this performance is the preliminary phase, during which the performers must familiarize themselves with the venue, its acoustics, and its emotional atmosphere. This helps them understand how these elements will interact with their voices. Special attention is given to the singers’ need to adjust their phonation and vocal emission, in order to adapt to the characteristics of the space and effectively convey their emotional message to the audience. The audience, in turn, is required to remain still in the benches, as dictated by the sacred nature of the venue.

From this initial study of the venue, the performers develop a precise spatial arrangement for each member in every piece to maximize the emotional impact on the audience. The effect of voices emanating from all directions, including positions not typically used in liturgical practices, allows spectators to experience the sacred space in a novel way, distinct from the central sound source that traditionally characterizes liturgical ceremonies.

For this thesis, only one piece was considered and analyzed, as will be discussed in the following chapters. The piece in question is “Crucifige,” which depicts the moment of extreme anguish and suffering during the crucifixion of Jesus Christ. It is characterized by a strong emotional intensity, with a soloist who sings as if announcing a message of death, while the other four male singers mimic the shouts of the crowd demanding Christ’s crucifixion. At the back of the church, another soloist, embodying the desperate Virgin Mary, laments the agony of her son’s death. In this way, sorrow and suffering become a shared experience, facilitating a deeper emotional connection and, ultimately, a sense of catharsis.

The same musical piece was recorded in each of the four churches, as well as in the anechoic chamber of the Politecnico di Torino, in order to compare the different features of the vocal delivery and verify the influence of the acoustics of the venues on the singers’ performance.

After the performance in each church, the comments of the singers were collected. They reported the auditory features of the venue, how they perceived the interaction with their voices and the changes the had to apply on their phonation to adapt to the different acoustics. Their preference of performing was also given at the end of the tour.

2.3 Acoustical Parameters

Room acoustics parameters had been calculated from the recordings in each of the churches. The detailed measurement procedure is explained in [25], and followed the protocol specific for churches proposed in [26].
In each church, 9 to 11 measurement positions were defined, with the sound source placed on the main altar and the microphones positioned between the benches and in the lateral naves. **Figure 2.4** presents the floor plans of the churches, indicating the microphone positions for each venue. The red dot represents the sound source, while the blue dot marks the position of the SMA used to acquire the IR for the auralization. The 3OA recordings were carried out using a real-time Bidule patch, which performs a real-time conversion from the 19-channel format to a 16-channel Ambisonics file, applying a convolution filter provided in [28].
For each position in each church, three identical sweep stimuli were recorded. An iterative MATLAB routine was then developed to compute the three IRs for each position. For the scope of this thesis, only the parameters registered at the position marked with the blue dot were considered. That is because for our analysis, it is important to consider the influence of acoustics on the singers: so, the acoustical parameters recorded in these positions, which are the closest to the perfomance position of the singer during the recordings, are considered to be more representative of the feedback experienced, compared to an average value among various positions in the environment. Objective parameters were computed from the IR recordings of both the omnidirectional microphone and the SMA.
These parameters are:

- Early Decay Time (EDT) and Reverberation Time (T_{20})
- Clarity (C_{80})
- Definition (D_{50})
- Center Time (T_s)
- Tonal Color (BR)
- Treble Ratio (TR)
- Inter-Aural Cross-correlation Coefficients (IACC$_{early}$ and IACC$_{late}$)

The acoustic parameters were computed using a MATLAB script that made use of the `ITA-Toolbox` library [29, 30]. To obtain a single value, an arithmetical mean of the values in the octave bands between 500 Hz and 1000 Hz was computed, except for IACC values, which were averaged on the octave bands from 125 Hz to 4000Hz as in [22].

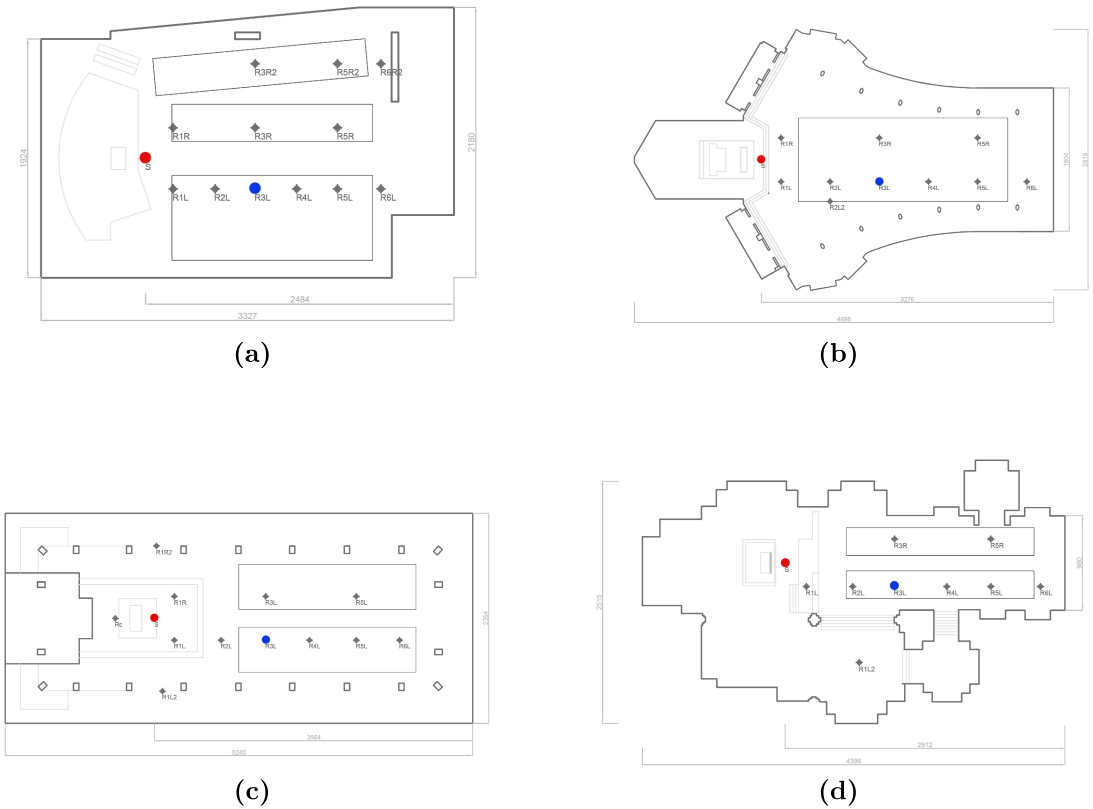


Figure 2.4: View of the interior of the churches: a) Church 1: Santa Gianna Beretta Molla; b) Church 2: San Giovanni Battista alla Creta; c) Church 3: San Nicolao della Flue; d) Church 4: Santi Giovanni Battista and Paolo.

2.3.1 Reverberation Time

The Reverberation Time (RT) is defined in [31] as the time required for a sound to decay after the sound source has stopped. It is influenced by the room's volume as well as the absorptive and diffusive properties of the walls, floor, and ceiling.
The RT is measured from the IR as the time it takes for the sound pressure level to decrease by a specified amount. These decay thresholds define different parameters used to describe this phenomenon: 10 dB for EDT, 20 dB for T_{20}, and 30 dB for T_{30}.
Specifically, T_{20} is calculated by tripling the time it takes for the level to drop from -5 dB to -25 dB SPL, while T_{30} is obtained by doubling the time for a decrease from -5 dB to -35 dB SPL. On the other hand, the EDT is determined using a linear regression of the first 10 dB of decay, with the resulting slope extrapolated to estimate the time for a 60 dB decay [32, 33].
An example of this calculations is presented in **Figure 2.5**.

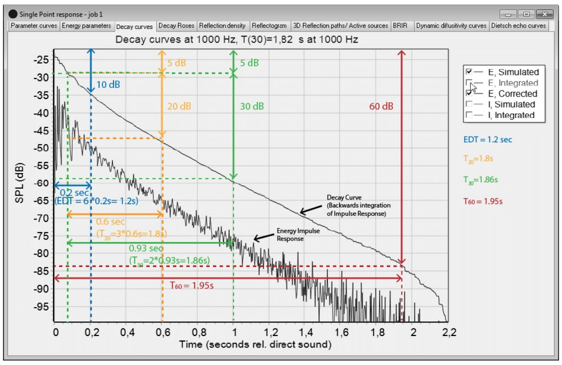


Figure 2.5: Calculation of Reverberation Time using Odeon [34] (image taken from [35])

In this study, both EDT and T_{20} were calculated, both in the actual churches and in the auralized environments.

2.3.2 Clarity

Clarity is a parameter that quantifies the quality of speech transmission from a source to listeners, taking into account both the direct sound and the reflections caused by the room acoustics. It is strongly influenced by the reverberation characteristics of the environment, which can negatively impact intelligibility.
For a theoretical impulsive signal, Clarity is typically quantified as the ratio of the energy of the sound arriving before and after a specified time threshold, leading to different metrics. For example, C_{50} is defined as the ratio between the energy arriving within the first 50 ms and the energy arriving afterward, whereas for C_{80}

the threshold is set at 80 ms.
Clarity is usually calculated through **Formula 2.1**.

$$C_{t_e} = 10\log_{10}\frac{\int_0^{t_e} h^2(t)\,dt}{\int_{t_e}^{\infty} h^2(t)\,dt} \tag{2.1}$$

In this definition, t_e stands for the time threshold to discriminate the early and the late sound (50 or 80 ms, as described before), while $h(t)$ is the room's IR.
The choice of whether to quantify clarity using C_{50} or C_{80} depends on the nature of the sound source being analyzed. While C_{50} is typically more suitable for speech-intended environments, C_{80} is more relevant as a parameter for music-dedicated spaces. For this reason, the latter will be considered in this study.

2.3.3 Definition

Definition is an alternative to Clarity in describing the quality of sound perceived by listeners. While Clarity is defined as the energy ratio between early and late arriving sound, Definition describes the ratio between early and total sound.
Similar to Clarity, different Definition metrics exist depending on the chosen time interval for early sound: D_{50} is used when the threshold is set to 50 ms, whereas D_{80} applies when the threshold is 80 ms. Definition is usually calculated according to **Formula 2.2**.

$$D_{t_e} = 10\log_{10}\frac{\int_0^{t_e} h^2(t)\,dt}{\int_0^{\infty} h^2(t)\,dt} \tag{2.2}$$

In this equation, t_e represents the time threshold chosen, while $h(t)$ denotes the room's IR. For this thesis, the time threshold of 50 ms, so the values corresponding to D_{50}, were computed.

2.3.4 Center Time

The Center Time (T_s) is commonly defined as the center of gravity of the squared IR of an acoustic environment. Analogous to the center of gravity in relation to the mass of a solid, T_s represents a weighted average of the energy distribution of the sound over time. It has been proposed as an alternative metric for assessing clarity [36].

$$T_s = \frac{\int_0^{\infty} t\cdot h^2(t)\,dt}{\int_0^{\infty} h^2(t)\,dt} \tag{2.3}$$

2.3.5 Tonal Color and Treble Ratio

Tonal Color is a parameter that characterizes the behavior of an acoustic environment in terms of energy distribution and timbral balance.
This measure is not yet ISO-standardized, but it has been proposed as a concept to characterize the tonal properties of musical spaces by *Gade et al.* [37]. However, to better account for human vocal characteristics, the frequency ranges used in this thesis to compute these ratios have been slightly adjusted compared to the original formulation.
Tonal color will be described through two parameters: the Bass Ratio (BR) and the Treble Ratio (TR). These parameters are particularly relevant as they quantify the spectral balance by comparing the energy in the mid-frequency range (centered at 500 Hz and 1000 Hz octave bands) with that in the low and high-frequency ranges. Specifically, the low-frequency range includes the 125 Hz and 250 Hz centered octave bands, while the high-frequency range comprises the 2000 Hz and 4000 Hz centered octave bands. Notably, BR has been shown to be strongly correlated with Room Timbre, which describes how an environment influences the perception of a sound's spectral components [24].
These parameters are computed as in **Formula** (2.4) **and** (2.5).

$$BR = \frac{RT_{125Hz} + RT_{250Hz}}{RT_{500Hz} + RT_{1000Hz}} \tag{2.4}$$

$$TR = \frac{RT_{2000Hz} + RT_{4000Hz}}{RT_{500Hz} + RT_{1000Hz}} \tag{2.5}$$

The RT in these formulae can be replaced with different definitions of Reverberation Time, as discussed in 2.3.1, depending on the specific analysis requirements.
For this study, these parameters were computed using the omni-directional IRs of each church through the `ita_roomacoustics_tonal_color()` function in MATLAB. This function, part of the `ita_roomacoustics` package [29], includes a parameter that allows selecting the specific definition of RT to be used in the computation from a set of standard options, including EDT, T_{20}, T_{30}, and T_{60}.
In this work, the timbral color introduced by the room response has been assumed to be strongly correlated with the perception of reverberance. For this reason, EDT was chosen, as it is widely recognized for its strong correlation with the subjective perception of reverberance.

2.3.6 Inter-Aural Cross-correlation Coefficient

The Inter-Aural Cross-correlation Coefficient (IACC) was computed from the 3OA IRs of each room. This parameter indicates the correlation between the signals

reaching the two ears. Its value ranges from -1 to 1: a value close to -1 indicates that the signals at both ears are identical but completely out of phase, while values approaching 0 suggest a higher degree of decorrelation between the signals.
Since the correlation between signals strongly depends on the propagation time from the emitter to the receiver, both Inter-Aural Time Difference (ITD) and Inter-Aural Intensity Difference (IID) are highly relevant. Furthermore, in a reverberant space, which introduces multiple secondary reflected paths, the directional differences in the arrival of these reflections also influence the resulting signals reaching the two ears. All these considerations are inherently subjective, as these parameters are difficult to model and are influenced by the individual characteristics of the listener's head and their unique Head-Related Transfer Function (HRTF).
This parameter provides crucial informations about the binaural, head-related similarity of the signals reaching the listener's ears and has been shown to be associated with the listener's perception of envelopment ([22], [38]).
The IACC is calculated (as described in [39]) starting from the Inter-Aural Cross-correlation Function (IACF), which is defined in **Formula 2.6**.

$$IACF_{t_1,t_2}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t) \cdot p_r(t+\tau)dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t)dt \cdot \int_{t_1}^{t_2} p_r^2(t)dt}} \tag{2.6}$$

In this formula, $p_l(t)$ and $p_r(t)$ represent the IR at the left and at the right ears, respectively, with τ denoting the time delay, and t_1 and t_2 as the integration limits within the signal's time duration.
The IACC is then calculated as shown in **Formula 2.7**:

$$IACC_{t_1,t_2} = max(|IACF_{t_1,t_2}(\tau)|) \tag{2.7}$$

for τ values ranging from -1 ms and 1 ms.
This parameter can be measured through direct calculation, but a binaural recording system is required to acquire the IR at the ear positions. An artificial head simulator can be used for this purpose. However, in this study, we opted to calculate these values from the 3OA IRs of each church. These responses were binaurally decoded using MATLAB, employing a standard HRTF. The resulting binaural signal was then processed using the `ita_roomacoustics_IACC()` function [29], which allows the calculation of the early, late, or full-time IACC, based on the definitions outlined in [31]. These three labels correspond to different pairs of integration limits in Formula 2.6.
The IACC values were obtained by averaging the values across octave bands from 125 Hz to 4000 Hz, as recommended in [22] and [23]. For each of the three parameters, both the mean and standard deviation were computed at the same position in each church.

2.4 Auralization System in the ASL

To explore the influence of reverberant spaces on human vocal comfort and performance, a subjective test protocol was designed. This was made possible by the Audio Space Lab (ASL), a laboratory at the Politecnico di Torino frequently used for experiments related to speech intelligibility and sound directivity. The laboratory is equipped with an Ambisonics reproduction system, comprising 16 loudspeakers (Genelec 8030B) and 2 subwoofers (Genelec 8351A) for low frequencies. Each piece of hardware is controlled by a set of two audio interfaces: one Roland OCTA-CAPTURE, used for input control, and one ANTELOPE Orion 32, which is fully engaged in managing the loudspeaker array. To minimize the latency and ensure the most real-time response of the system, the buffer sizes of both audio interfaces were set to 32 samples for the OCTA-CAPTURE and 64 samples for the ANTELOPE Orion. The room has been treated to reduce room modes and ensure a certain level of sound insulation. The following **Figure 2.6** shows a view of the laboratory. Details reguarding the calibration and validation procedures of the laboratory may be found in previous literature [40, 41].

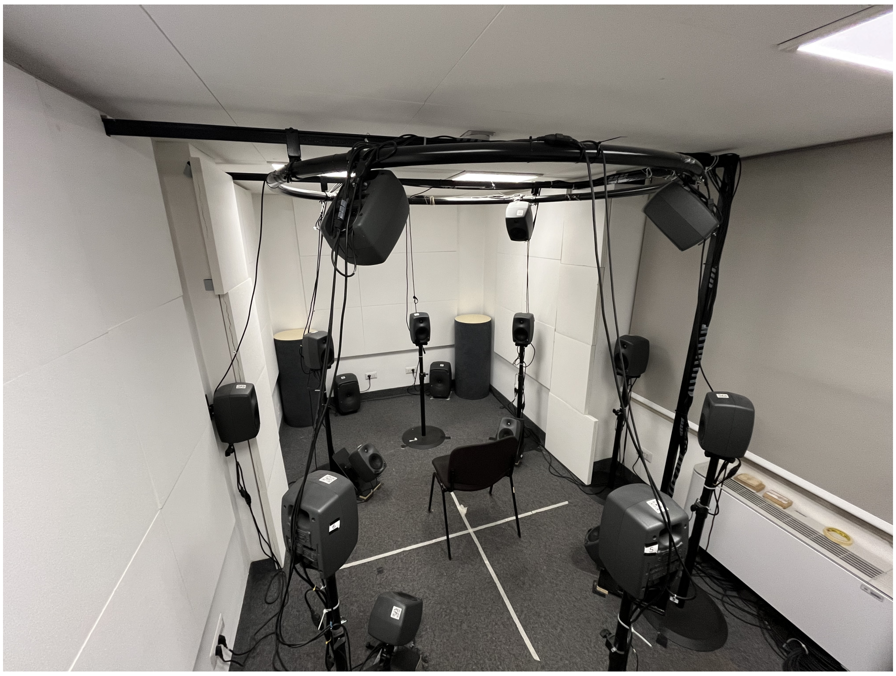


Figure 2.6: A view of the Audio Space Lab

2.5 Acoustics Parameters in the Audio Space Lab

To describe the acoustic features of the convulted environment created through the loudpseakers array in the ASL and reaching the performers' ears, both a group of omnidirectional parameters and binaural acoustics parameters were calculated from the (IR) of each auralized environment in the ASL.
Both EDT and T_{20} were computed to compare the reverberant tails obtained in the four auralized situations. Binaural parameters were derived by analyzing the OBRIRs of each auralized environment, measured using an artificial head simulator (Head Acoustics HMS II.3 LN HEC, see **Figure 2.7**) with loudspeakers and microphones to simulate both human vocal and hearing apparatus. The OBRIRs were then filtered across octave bands from 125 to 4000 Hz, as this range encompasses the majority of frequency features of human emission. The resulting parameters were obtained for each octave band within the 125 to 4000 Hz range. To obtain a single value, an arithmetic mean was calculated by averaging the values from the 500 Hz and 2000 Hz octave bands, as in [17].
The results for the signals from the two ears were then averaged to obtain a unqiue value. Also, the standard deviation within the values from the three OBRIRs was taken into account.

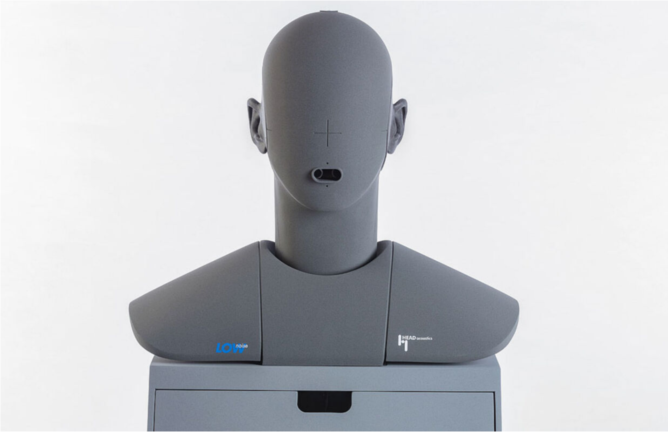


Figure 2.7: HEAD Acoustics - HMS II.3 LN HEC artificial head

2.5.1 Room Gain

Room Gain (G_{RG}) was defined by *Brunskog et al.* as the amplification degree applied by the effect of room acoustics on a talker's voice, which can also be perceived by the talkers themselves [15]. It can also be defined as the ratio of the energy of the direct sound arriving at the speaker's ears, compared to the total

energy of the sound, considering an impulsive emission [42].
This theoretical interpretations lead to the definition of G_{RG} as the ratio between the total energy of the OBRIR and the energy level of the direct sound, considered to be included in the first 5 ms of the signal. The G_{RG} is expressed in decibels. The energy levels are calculated as follows:

$$L_E = 10 \log_{10} \frac{\int_0^\infty h^2(t)dt}{E_0} \tag{2.8}$$

$$L_D = 10 \log_{10} \frac{\int_0^{5ms} h^2(t)dt}{E_0} \tag{2.9}$$

In this equations, $h(t)$ is the OBRIR calculated using an artificial dummy head, with the sound being emitted from the mouth and recorded at the ears. E_0 is a reference energy level, which could be chosen arbitrarily. The G_{RG} is then computed as the difference between these two values, as in **Formula 2.10**

$$G_{RG} = L_E - L_D \tag{2.10}$$

2.5.2 Vocal Support

The Vocal Support (ST$_V$) was defined as the difference between the reflected and the direct sound energy levels, while considering a talker emitting sound and receiving it at their ears. It was defined by *Pelegrín-García et al.* as an alternative to Room Gain [43].
Similar to the calculation for the G_{RG}, also for the ST$_V$ a time threshold of 5 ms is considered to distinguish between the direct and the reflected sound. The energy levels are calculated as follows:

$$L_D = 10 \log_{10} \frac{\int_0^{5ms} h^2(t)dt}{E_0} \tag{2.11}$$

$$L_R = 10 \log_{10} \frac{\int_{5ms}^{\infty} h^2(t)dt}{E_0} \tag{2.12}$$

In this equations, $h(t)$ is the OBRIR calculated using an artificial dummy head, with the sound being emitted from the mouth and recorded at the ears. E_0 is a arbitrary reference energy level.
The ST$_V$ is finally computed as in **Formula 2.13**

$$ST_v = L_R - L_D \tag{2.13}$$

Assuming that the sum of direct and reflected sound is equal to the total energy of the sound, which may be an approximation [42] due to the nature of the calculations themselves, a relationship between the G_{RG} and the ST$_V$ can be established, as in **Formula 2.14**:

$$G_{RG} \approx 10 \cdot \log_{10} \left(10^{\frac{ST_V}{10}} + 1\right) \tag{2.14}$$

2.5.3 Decay Time Mouth-to-Ears

The Decay Time 40 Mouth-to-Ears ($\mathrm{DT}_{40,ME}$) has been defined in [16] as the time it takes to the reversed integrated energy curve of an IR sound to decay by 60 dB, after the direct sound has been emitted by the talker and received at the ears (from which the subscript "*ME*" comes from). In the specific case of this thesis, the IR used will be the OBRIR recorded at the ears of the dummy head in the ASL, while the emitted sound is real-time auralized be menas of the convolution system developed.
This parameter is calculated using the following **Formula 2.15**.

$$EDT_{40,ME} = \frac{60}{X} \cdot (t_{-X} - t_0) \tag{2.15}$$

In this equation, t_{-X} is the time needed for the sound intensity to drop to a value of X dB lower than the initial level, starting from the initial time t_0.
As for RT (see 2.3.1), this value is calculated considering a linear projection of the initial decay of X (in our case, this decay amount was set to 40 dB). It is worth noticing that by changing this initial decay value to any value of choice, it would be possible to define various decay times, which may better adapt other specific applications.
This parameter is a significant value that can describe the energy of the reverberant tail that arrives at the talker's ears and may influence his perception of the environment. In our study, this is of particular interest because of the nature of the auralized environments, which are churches with different acoustics and may affect the singers delivery, as reported in the previous chapters.

2.6 Voice Parameters

In order to analyze the variations of the vocal delivery characteristics of the singers in the various auralized environments, a series of parameters were used. These indexes have been reported to be descriptive of diverse features of the human voice. It is important to note that the majority of these parameters have been scarsely applied to the spefic context of the singing voice, and via specific testing procedures; so the results that will be presented in the following chapters are scarsely comparable to others reported in previous literature. Nonetheless, this procedures may be a suggestion for future development and research in the field of singing analysis through the recording of a musical piece.

2.6.1 Sound Pressure Level

Measurements on the sound emission intensity level were performed to verify a correlation between vocal comfort and reverberation characteristics of the various environments. To do so, the Sound Pressure Level SPL (dB) was calculated on the recordings. The calibration procedure will be deeply described later on, but after that procedure it was possible for us to compute the SPL level using the same calibration signal intensity provided by the recording of the calibrator output on a reference microphone (in our case, the calibrator was the B&K 4231, while the reference microphone was the one provided with the NTi Audio XL2 level meter). For this reasons, the values presented are pretty high, as they were captured close to the signers' mouths.
The level was computed using the following **Formula 2.16**.

$$SPL = 20 \cdot \log_{10} \left(\frac{rms(x(t))}{20 \cdot 10^{-6} \cdot rms(x_{cal}(t))} \right) (dB) \tag{2.16}$$

In this equation, the function $rms(\cdot)$ represents the Root Mean Square of the array provided in input, $x(t)$ is the signal recorded with the headworn microphone, and $x_{cal}(t)$ is the calibration signal.
It is important to note that, because of the recording procedure adopted in the churches, the results obtained during the recordings of the professional solo singer in the churches were not acceptable, and will not be considered during the data analysis.

2.6.2 Time Dose

Another parameter taken into account was the ratio of the voiced segments in the recordings compared to the total duration [44, 45]. This parameter, referred to as Phonation Time or Time Dose (TD), is defined as the ratio of the number of voiced frames to the total number of frames, as expressed in **Formula 2.17**.

$$TD = \frac{\#voicedFrames}{\#Frames} \tag{2.17}$$

For this parameter, the signal was subdivided in frames of 1024 samples (which corresponds to a temporal length of 21.3 ms at a sample rate of 48,000 Hz). To discriminate the voiced frames from the unvoiced ones, a threshold was chosen, corresponding to half the value of the Root Mean Square of the entire recording. Then the Root Mean Square was computed for each single frame: if the value is lower than the threshold, the frame is considered as an unvoiced frame.

With this calculation, informations about the length of unvoiced frames are also derived, such as the distribution of the length of successive voiced frames and successive unvoiced (named as "*pause*" frames).

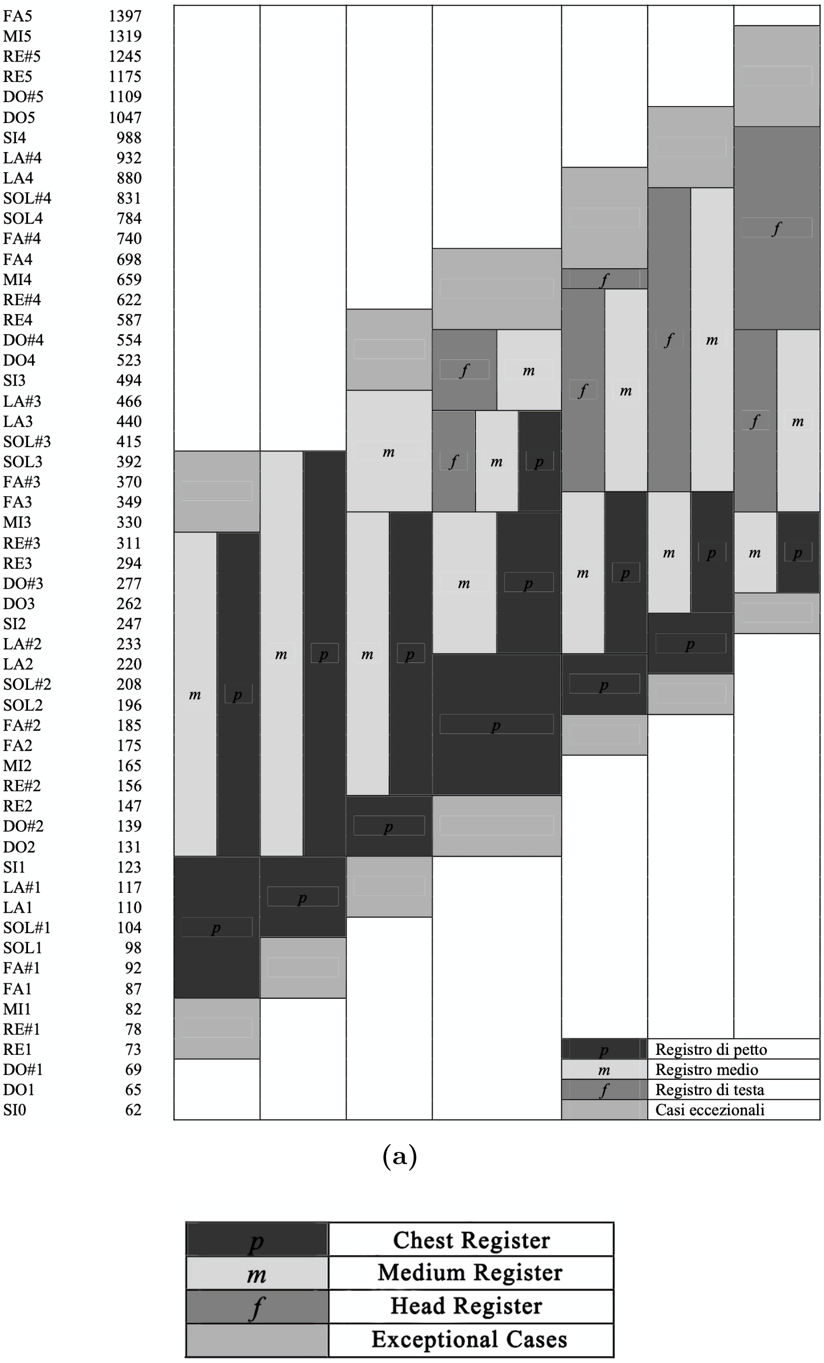
It is worth noticing that this parameter was only computed for the auralization recordings, since the nature of the recordings in the churches was not ideal to compute this measurements, as reverberant tails from the venues makes it impossible to correctly distinguish voiced from unvoiced frames using a threshold. Therefore, these parameters on the soloist recordings were not considered in the present work.

2.6.3 Fundamental Frequency

The human vocal emission is largely conditioned by many factors, such as the vibration of the vocal folds, the control of the air flow by the diaphragm and the resonances of the vocal tract: all these factors introduce a peculiar filtering on the signal spectral envelopment (which are called the Formants), varying from person to person.

While considering a singing task, the human voice is modulated by the performer in order to produce a complex combination of harmonics. Professional singers are trained to control every aspect of their emission, in order to exploit the full potential of their voice with the lower effort. Various techniques are used to achieve this result, depending on the musical style and subjective characteristics of each singer.

The Fundamental Frequency (F_0) is commonly associated with the lowest harmonic of the voice, corresponding to the vibration of the vocal folds as filtered through the vocal tract. This parameter is typically linked to the concept of *pitch*, a perceptual attribute related to the perceived musical note of a sound. However, pitch perception does not always correspond exactly to the actual F_0 of the emitted sound. F_0 is strongly correlated to the singer's vocal range, which is the interval of frequency (hence, notes) a singer can reach while performing. The traditional, European classification of vocal ranges is depicted in **Figure 2.8**.



(a)

(b)

Figure 2.8: Vocal ranges classification. From left to right: Bass, Bariton, Tenor, Alto, Mezzo-Soprano, Soprano. [46]

In the figure, vocal ranges are presented in an ascending order from the left to the right. Each vocal range correspond to a specific classification of vocal types, based on the singer's *tessitura* and registers. The legend corresponding to the registers depicted in Figure 2.8a has been translated from the Italian and reported in Figure 2.8b.
During the recording session of the subjective experimentation in the ASL, amateur singers were provided with a fixed reference note, selected according to their vocal range, to ensure they could perform comfortably. This approach aimed to standardize the starting pitch for individuals with similar vocal ranges and to account for the exact note assigned to each participant during data analysis. Consequently, the data analysis phase focused on evaluating the variance of the fundamental frequency relative to the reference note given by the operator before each recording. In this thesis, the F_0 was obtained by means of a MATLAB script that performed an auto-correlation calculation on frames of fixed length (1024 samples, *i.e.* 21.3 ms at sample rate 48,000 Hz).

2.6.4 Long-Term Average Spectrum

Many insights into the characteristics of vocal production can be derived from spectral analysis. To obtain a general overview of the entire recording in each environment, the Long-Term Average Spectrum (LTAS) is used. This is a plot of the average spectrum across multiple frames. In this work, the frames contained 1024 samples, with a sampling rate of 48,000 Hz, resulting in a frame length of 21.3 ms. By increasing the length of the frames, a higher frequency resolution is achieved, although this comes at the cost of considering a longer time window simultaneously. The spectra can be obtained using **Formula 2.18**

$$S_i(f) = 10 \cdot log_{10}(|FFT\,(x_i(t) \cdot w(t))|^2) \tag{2.18}$$

where the i-th temporal frame is weighted using a Hamming window *w(t)* before applying the Fourier transform. Then, the result is inserted into the logaritmic operator to convert to a logaritmic scale. Finally, to obtain the average spectrum between each frame, a mathematical mean is computed as in **Formula 2.19**

$$LTAS(f) = \frac{1}{N}\sum_{i=1}^{N}(S_i(f)) \tag{2.19}$$

where N is the number of temporal frames in which the original audio file has been subdivided.
The LTAS is useful to analyze the spectral features of an audio excerpt. By using this instrument, it is possible to consider a general glance upon a mean of the distribution of the energy between the different frequencies across the whole duration

of a signal, instead of focusing on single windows, which can only be representative of few milliseconds of the musical piece.

2.6.5 LTAS Features

Many pieces of information can be derived from the LTAS. Regarding the singing aspects, is important to study the energy in the frequency range between 2000 and 4000 Hz. Many previous studies have shown that an increase of the spectral envelope in this area is linked to a more pleasant and musical sung result, probably due to the presence of Singer's Format, which is a phonatory modulation in the voice, mostly visible in trained singers.
Firstly, the energy in dB in three bands is calculated, considering the frequency ranges between 0 Hz and 1000 Hz (*B0*), between 1000 Hz and 4000 Hz (*B1*) and between 4000 Hz and 9000 Hz (*B2*) [47].

The energy in each range is computed as in Formula 2.20

$$B = 10 \cdot \log_{10}(\sum_{i=1}^{n} \left(10^{\frac{LTAS(n)}{10}}\right)) \text{ (dB)} \tag{2.20}$$

where n is the length of the vector of the LTAS in the selected frequency range. Then, the differences are computed as in **Formula 2.21** and **2.22**:

$$\Delta B_{low} = B_1 - B_0(\text{dB}) \tag{2.21}$$

$$\Delta B_{high} = B_2 - B_0(\text{dB}) \tag{2.22}$$

Lã et al. [48] analyzed the singing voice features focusing on Fado-Cançäo, a traditional Portuguese musical style recognized by UNESCO as "World's Intangible Cultural Heritage".
In particular, the Formant Cluster Prominence (FCP) was used as a index of prominence of harmonics in the range from 2000 Hz and 4000 Hz. Unlike other indexes that may be used to describe the spectral prominence in this area compared to the intensity of F_0, the FCP is computed by means of **Formula 2.23**:

$$FCP = max(LTAS(f)_{f \in [2000;4000]Hz} - reg(f)_{f \in [1000;5000]Hz}) \text{ (dB)} \tag{2.23}$$

where the regression line *reg(f)* of the mean LTAS in the range between 1000 Hz and 5000 Hz is subtracted from the value of the mean LTAS in the range of frequencies between 2000 Hz and 4000 Hz.
By means of this calculation, FCP is calculated from the mean LTAS and can be considered as an indicator of prominence of the cluster of harmonics in the range from 2000 Hz and 4000 Hz, but its value is compensated for the spectral slope, indipendently from task, singers' skill and style of the sung piece.

2.6.6 Singing Power Ratio

The Singing Power Ratio (SPR) was defined in [49] as an indicator of quality of singing, rather than an indicator of the presence or absence of the Singer's Formant. It is also reported to be correlated to the "ringing" quality of singing voice, and to be generally greater in professional singers rather than non-professionals [50].
The SPR is defined as the difference of the peaks of the spectrum in the frequency range from 0 Hz to 2000 Hz, and the frequency range from 2000 Hz to 4000 Hz, as depicted in the following **Formula 2.24**.

$$SPR = max\left(S\left(f\right)\right)_{f\in[2000;4000]Hz} - max(S(f))_{f\in[0;2000]Hz} \tag{2.24}$$

In this equation, $s(f)$ is the spectrum of the signal, obtained via Fast Fourier Transform in frames of fixed length of 1024 samples (*i.e.* 21.3 ms at a sample rate of 48,000 Hz). To obtain a more precise result, an overlapping of half the length of the frame is used. Then, an arithmetical mean between all the values across the performance is computed.

2.6.7 Cepstral Peak Prominence Smoothed

In the past, some research has shown a link between the cepstral features of a signal and the vocal effort related to the sound emission. The term *Cepstrum* itself gives a hint about the nature of this transform: it is obtained by computing the inverse Fourier transform of the logarithm of the spectrum amplitude of a signal, as shown in Formula (2.25)

$$C_p = 20 \cdot \log_{10}\left|\mathcal{F}^{-1}\left\{20 \cdot \log_{10}\left(|\mathcal{F}\{s(t)\}|\right)\right\}\right| \tag{2.25}$$

where C$_p$ is the Cepstrum vector, $\mathcal{F}$ is the Fourier Transform and *s(t)* is the original signal in time domain. Through the Cepstral Analysis of a signal, it is possible to verify the presence of periodic elements of the spectra, such as harmonics, reflections, or echoes. In the *Quefrency* domain, these elements will be displayed as isolated peaks.
In particular, in this work the *Cepstral Peak Prominence Smoothed* (CPPS) was considered and computed using a MATLAB script. Some precautions were needed: in fact, the original signal was resampled to a sampling rate of 22,050 Hz. Then, the spectrum of the signal was calculated considering a time window of 1024 frames. To reduce the influence of noise and reduce the flactuations in the signal, a temporal smoothing was applied to the signal, averaging over 7 temporal frames.
Afterwards, a smoothing in the cepstral domain was also applied: due to the nature of the calculations used to computed the cepstrum, it can be described as a sequence of discrete bins of fixed width in the cepstral domain, with a nominal

value for each bin. So, to attenuate the fluctuations in this domain, an averaging for each bin was applied, considering 3 bins before and 3 bins after the current bin.

The Cepstral Peak Prominence is a measure of the heigth of the peak in the cepstral domain. It is calculated as the difference between the maximum of the cepstrum and a regression line, which was derived considering quefrency values between 2 ms and 16 ms. The boundaries were chosen to be the reciprocal of 60 Hz and 500 Hz, respectively, which was considered as a range of values wide enough to include the F_0 of the singing performance of the majority of the vocal registers.
The low quefrency part of the cepstrum is mathematically associated with the envelopment of the spectrum, which may present a slow periodicity due to the harmonic nature of the signals analyzed. Therefore, the regression line was calculated considering the quefrency range from 1 ms till the end.
The CPPS is then calculated for each frame by computing the difference between the maximum value of the cepstrum - in the quefrency range from 60 Hz to 500 Hz - and the regression line. An example of this calculus on a single frame is shown in the following **Figure 2.9**.

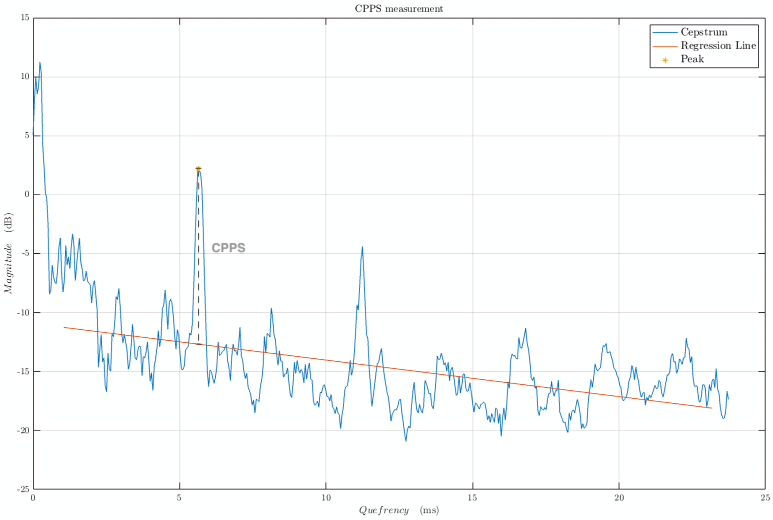


Figure 2.9: CPPS evaluation on a frame

The CPPS values can be displayed as the distribution of values in a distribution graph. Much literature reported a relation between the distribution of CPPS values and the health status of the voice of normophonic and dysphonic subjects [51]. Nevertheless, it is yet to be determined if this parameter can be used to describe meaningful informations about the singing performance of a subject [52].

Chapter 3

Auralization System Setup & Calibration

In this chapter, the various steps of the auralization system setup and calibration are presented.

First of all, the pre-processing on the IRs from the original four churches is discussed. Then, the measurement procedure of the latency introduced by the system is presented. Lastly, the real-time auralization system design and its calibration are presented. For the calibration procedure, an omni-directional microphone connected to a sound level meter (NTi Audio XL2) was used, along with an artificial head simulator by HEAD Acoustics, specifically the HMS II.3 LN HEC.

3.1 Pre-processing

Prior to the setup of the auralization engine, some pre-processing was needed to modify the IRs. All the following steps were performed using a MATLAB script.
First, during the measurement sessions, three 3OA IRs were recorded for each church. Since selecting one arbitrarily would be subjective, it was decided to average them. This approach ensures that the resulting IRs closely approximate the actual acoustic response while reducing the impact of noise that may affect individual recordings.
Next, the IRs of each environment were normalized. The highest absolute value among the 16 channels of each IR was identified and used to normalize all channels of the same IR.
Finally, an additional modification is required to adjust the auralized content reproduced by the system. Generally speaking, an IR can be subdivided in sections, as shown in **Figure 3.1**.

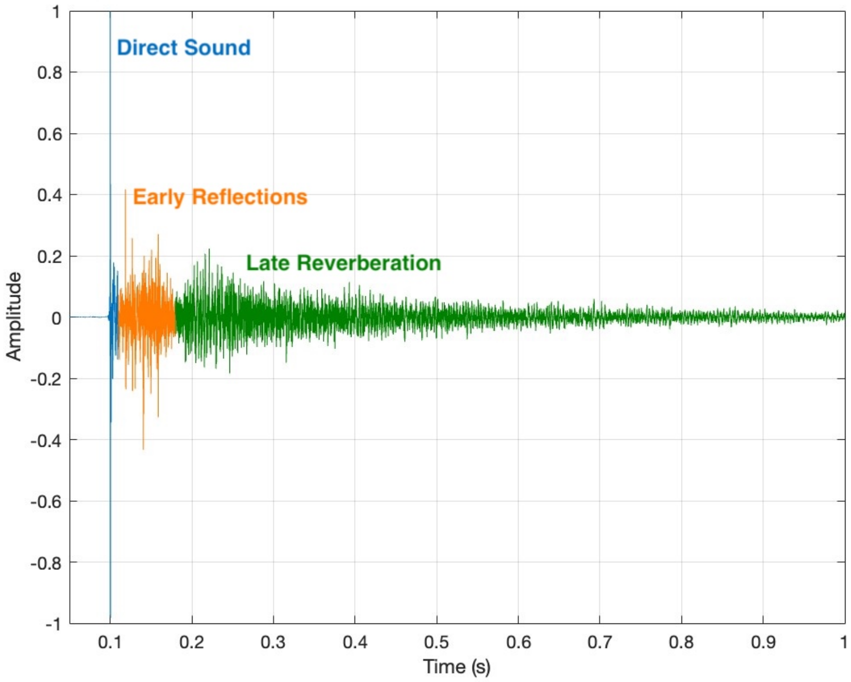


Figure 3.1: Structure of a room impulse response

The first section, which follows the higher peak in absolute value of the IR, is assumed to be the direct sound, which is the component that arrives first and with the highest energy. The whole direct sound is generally accepted to be contained within 0 ms to 10 ms from this peak [31]. The time delay from the source emission to the receiver perception is referred to as *flying time*, and depends on the time interval needed for the sound to travel along the direct path from the source to the receiver.

The following components in the IRs comprehend the first reflections, that fade into a diffuse field as long as later reflections from the reverberant environment arrive. These latter parts should be maintained, since these sections of the IR contribute the most in reconstructing the late diffuse field of the reverberant environment.

In the ASL, the arrival of the direct sound is granted by the direct path from the speaker's mouth to his own ear, so it should not be reproduced from the auralization loudspeakers. To exclude this first part, each channel of the Ambisonics IR was filtered to nullify the first 10 ms. A Tukey window-shaped filter, with a rise time of 2 ms (*i.e.* 96 samples at 48,000 Hz), is applied to each channel to avoid the presence of any steps, which may have bad consequences on the convolution result, such as auditory artifacts. An example of such a filter is shown in **Figure 3.2**.

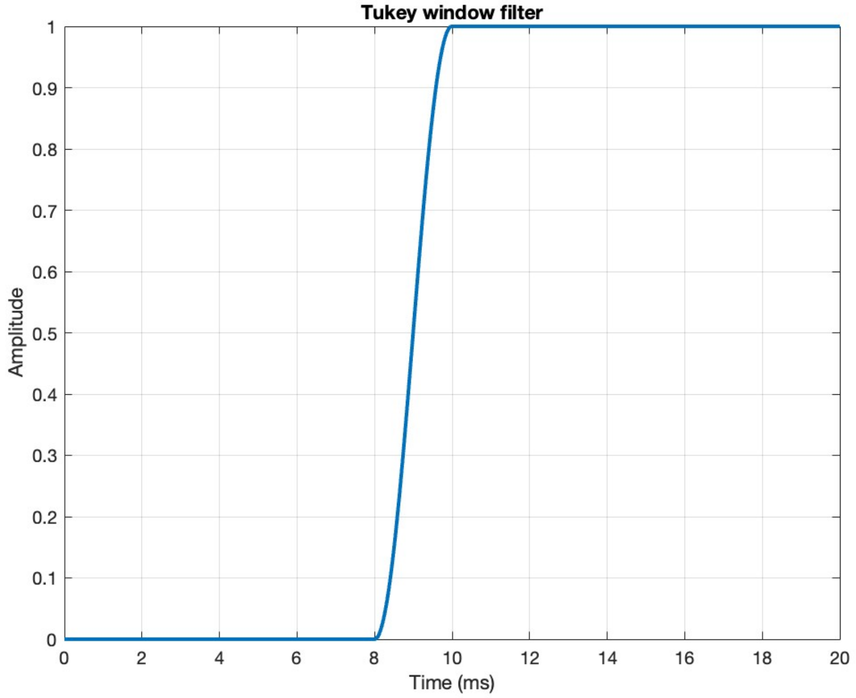


Figure 3.2: Filtering window used to remove the direct sound

The value of the filter has been set to 0 up to 8 ms after the arrival of the direct time, and rises to 1 within the rise time, so that the components beyond 10 ms are preserved. This grants that the loudspeakers do not reproduce the direct sound, while maintaining the later reflections and the reverberant tail. An example of the filtering applied to the IRs is shown in the following plot in **Figure 3.3**.

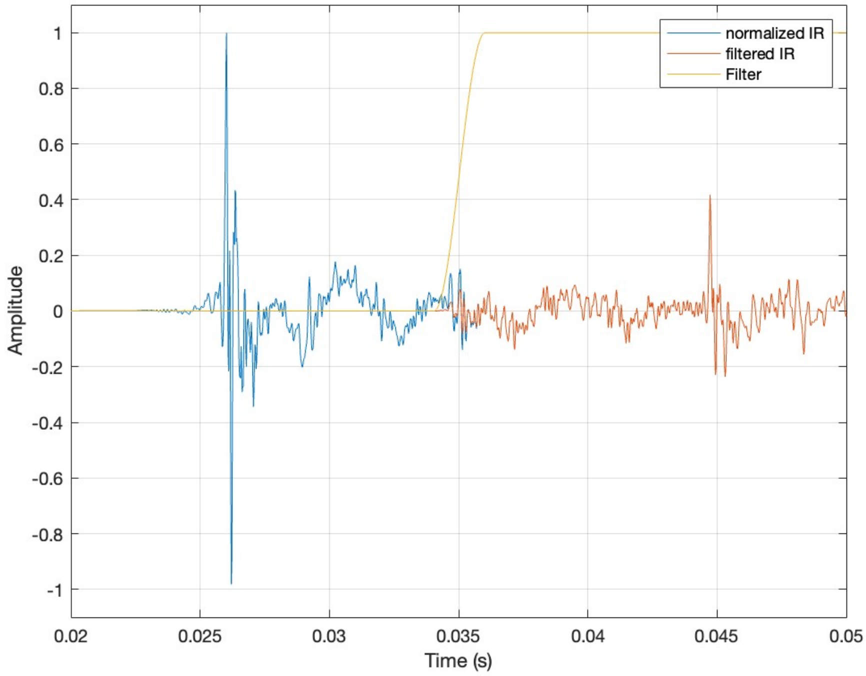


Figure 3.3: Example of the direct sound removal, using a properly time-shifted windowing

The system latency should also be taken into account. Since this parameter would introduce an unnatural delay in the arrival of the auralized sound, a number of zeroes had been removed from the beginning of each channel of every IR, as suggested in [9]. This number of samples is computed using the procedure depicted later in this chapter (see Section 3.2.1).

3.2 Auralization System: Design & Calibration

3.2.1 Latency Measurement

The latency introduced by the auralization system is a crucial factor, as users may perceive a loss of realism if the delay introduced during signal processing exceeds the threshold of perceptible difference. To measure latency, a dedicated convolution system was designed, closely resembling the one used during the auralization subjective experience.

The 3OA IR used in the convolution step was the one recorded in Church 1. The pre-processing step differed slightly from the one described in Section 3.1: the signal was normalized and averaged across the three iterations, then normalized to the peak value among all of the channels. However, the direct sound was not filtered out, as it is fundamental to the developed methodology.

In the Plogue Bidule modular audio software [53], a patch was designed, as illustrated in the block diagram in **Figure 3.4**.

In this block diagram, the various blocks represent different stages of sound processing, implemented directly within the Bidule patch. In particular,

- *Gin* and *Gout* are two gain nodes, set to prevent clipping conditions in the input and output stages, respectively;
- The **CONVOLVE** node is the Convolver node. Specifically, the X-MCFX Convolution Node [54] was used, with a first partition size of 64 bits and a maximum partition size of 8192 bits. The IR used for the convolution was the one from Church 1, averaged across the three iterations and normalized.
- The **AMBIX DECODER** and **EQ** nodes are designed to ensure accurate reproduction of the 3OA signal in the ASL loudspeaker array.
- The **Louspeakers Array** and the **Microphone** nodes are a schematized sum-up of the setup in the ASL, were a reference omnidirectional microphone (NTi Measurement microphone) is placed in the sweet spot of the loudspeakers array and connected to the audio interfaces of the laboratory. The input is then available for the recording in Bidule.
- The **Recorder1** and **Recorder2** nodes are two virtual recorders that capture the input sound and store it in a `.wav` file, using 32-bit floating point encoding. The two files were named *RecPlayer.wav* and *RecMic.wav*.

A sweep signal was generated using Adobe Audition software [55], with a sample rate of 48,000 Hz. The sweep duration was 500 ms, spanning a frequency range from 90 to 10000 Hz. Both the sweep and the inverse filter were generated and saved. The sweep was loaded using the `Player` node in Bidule and convolved

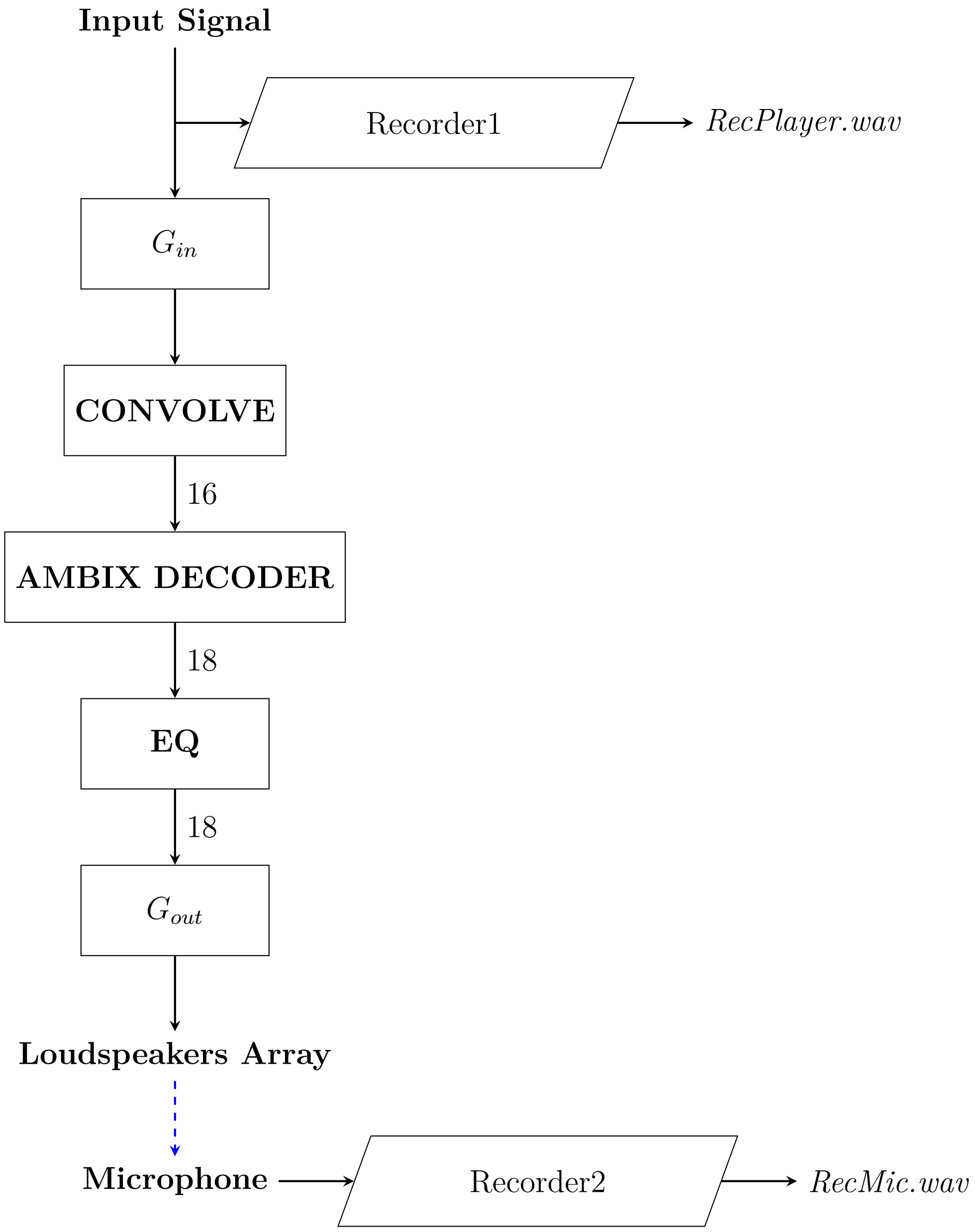


Figure 3.4: Block diagram of the Latency Measurement System

with the IR. The resulting 3OA signal was then decoded for reproduction in the loudspeaker array and equalized to ensure correct reproduction. The fidelity of the reproduction system of the ASL has been validated in other studies and is considered to be verified. For further details on this aspect, please refer to [41, 40]. The laboratory setup of this phase is shown in **Figure 3.5**.

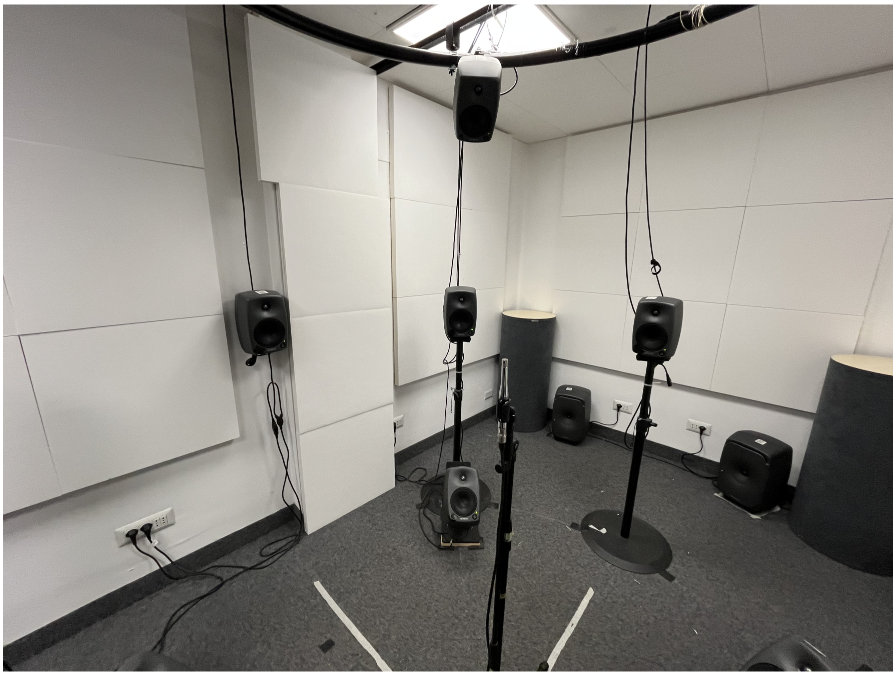


Figure 3.5: Setup for latency measurement

The convolved signal is reproduced by the ASL loudspeaker array and then sampled by the microphone, which is connected to the Roland OCTA-CAPTURE audio interface. The input signal is then available in Bidule. Both the played sweep and the input signal are recorded using two `Recorder` nodes in the Bidule patch, with their "Recording" status linked to the "Playing" status of the `Player` node. This ensured perfect synchronization between the start of the two audio files in output. A MATLAB script was then designed to perform the convolution of the audio files using the inverse filter generated previously. Through this convolution, it is possible to calculate the IR of the linear system that modifies the original signal. Both the recordings and the inverse filter were loaded into MATLAB using the `ita_read()` function (part of the `ITA_toolbox` package, [29]). This function creates an `ita_audio` object, which stores various information, such as the sampling rate and the array of audio samples.

In this specific case, the de-convolved signal extracted from the *RecPlayer.wav* file should tend toward a perfect impulse, as the sweep signal is recorded directly, without the intervention of any processing. On the other hand, the deconvolved signal extracted from the `RecMic.wav` file should exhibit a notably non-impulsive behavior, as it sampled the sweep after convolution with the IR and was then played through the loudspeaker array. A sample of the result from this step in reported in **Figure 3.6**. For the sake of clarity, in this plot the two curves have been normalized according to their respective maximum absolute values.

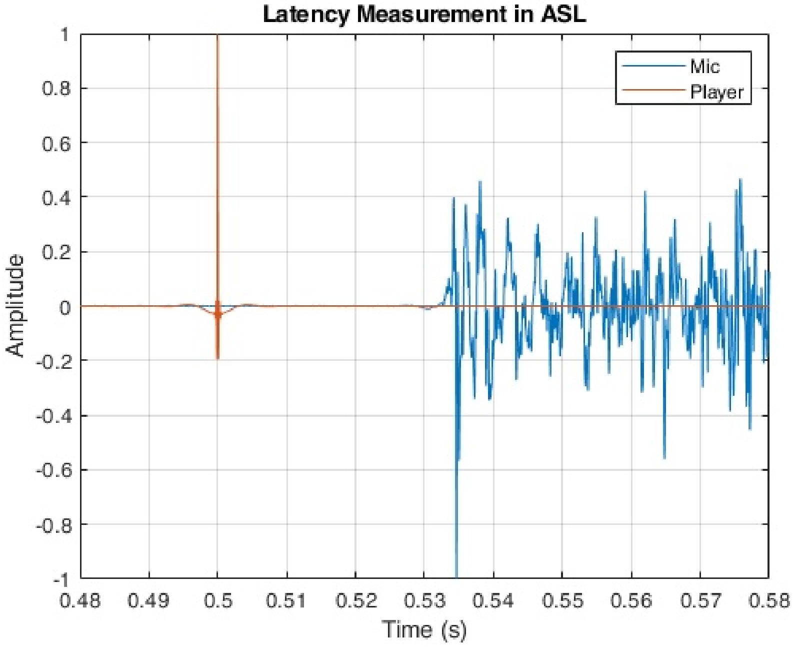


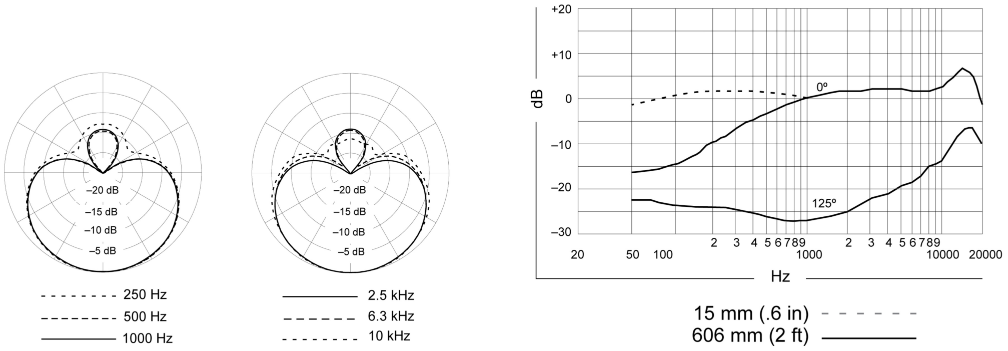
Figure 3.6: Plot of the IR from the two recordings

According to an assumption made in previous sections, the direct sound can be considered to be represented by the highest peak in absolute value within an IR. As can be seen in Figure 3.6, the two IRs clearly show a time difference regarding the direct sound arrival. This difference can be considered to be the value of the latency introduced by the system setup, which was measured to be 1662 samples (i.e. circa 30 ms at a sampling rate of 48,000 Hz). It is important to note that this value may vary according to the system's CPU load; however, these variations have been considered negligible for the purpose of this preliminary study.
It is also crucial to note that the value of this latency includes the flying time, which should not be accounted for in the latency computation, as it is an environment-dependent variable that should be maintained during the auralization process.
Therefore, by subtracting the value of the flying time (which was equivalent to

1282 samples in Church 1) from the initial value of 1662 samples, the resulting latency time was calculated to be 380 samples (i.e. circa 7.9 ms at a sampling rate of 48,000 Hz).
This value was used to reduce the number of leading zeros from each environment to compensate for the delay introduced by the realt-ime auralization system. To verify the accuracy of this compensation, the same procedure described in this section was repeated using the modified IR, from which the calculated amount of leading zeros had been trimmed. As expected, the difference between the two absolute value peaks in the `RecPlayer.wav` and `RecMic.wav` recordings corresponded to the flying time of Church 1, with a maximum variability of ±1.5 ms, which is practically imperceptible.

3.2.2 Real-Time Auralization Procedure

During the subjective experimental procedure, auralization was performed using a Bidule patch that executed a mathematical convolution between the input signal and the modified IRs of the churches. In this setup, the input signal from the singers' voices was captured using a head-worn microphone (SHURE WBH54) connected to the Roland OCTA-CAPTURE audio interface. The processed signal was then fed into the Bidule patch and subsequently reproduced through the spherical loudspeaker array, driven by the ANTELOPE Orion 32, a secondary audio interface.
The head-worn microphone selected for the experiment has a hypercardioid polar pattern and a fairly linear frequency response at short distances (both the polar pattern and the microphone's frequency response are shown in **Figure 3.7**).



(a) **(b)**

Figure 3.7: SHURE WBH54 typical polar pattern (a) and frequency response (b) (data taken from [56])

By using a hypercardioid microphone, it is possible to capture only the audio originating from the singer's mouth while attenuating sound from other directions. Due to the microphone's reduced sensitivity to off-axis sounds, the auralization from the loudspeaker array is significantly attenuated or not recorded at all. This was an important issue during the early design stages of this thesis work, as the feedback from the auralization through the input microphone would have been a destructive phenomenon.
Inside the Bidule patch, the signal was first convolved with the modified IR of a church and then passed through an Ambisonics decoder, which calculated the output signal that each loudspeaker should reproduce. A schematic representation of the system is shown in the following **Figure 3.8**.
Both the convolution and the Ambisonics decoding were performed using the X-MCFX Convolution node [54], with a first partition size of 64 bits and a maximum partition size of 8192 bits. For a detailed description of the Ambisonics calibration of the ASL please refer to [41].

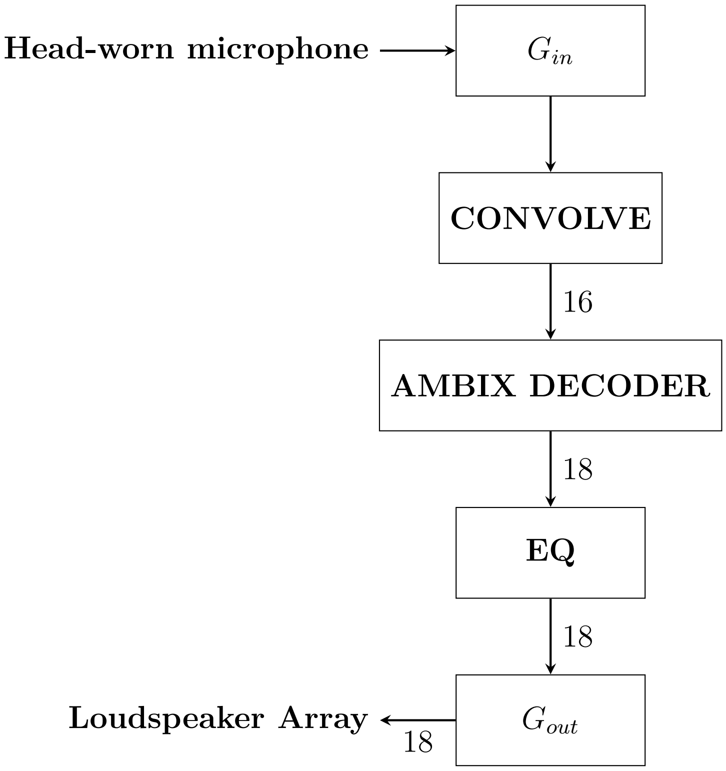


Figure 3.8: Block diagram of the Real-Time Auralization System

In the block diagram in Figure 3.8:

- G_{in} and G_{out} are two gain nodes, set to prevent clipping conditions in the input and output stages, respectively. G_{out} node underwent a specific calibration procedure, which will be discussed in detail in Section 3.2.3;
- The **CONVOLVE** node is the Convolver node, which perform a real-time convolution with the IRs derived from the editing of the ones from the four churches.
- The **AMBIX DECODER** and **EQ** nodes are designed to ensure accurate reproduction of the 3rd Order Ambisonic Signal in the ASL loudspeaker array.

Additionally, the weights positioned beside some of the arrows represent the number of audio signal channels passing from one node to the next. In particular, the convolution of the mono signal—sampled using the head-worn microphone and amplified in the G_{in} node—with a 3rd Order Ambisonics IR results in a 16-channel audio signal. Furthermore, this Ambisonics signal must be decoded into an 18-channel signal for reproduction through the system in the ASL.

3.2.3 Auralization Calibration

The system was calibrated to ensure that the same acoustic level reaches the user of the auralization engine.
Firstly, the head-worn microphone was calibrated to emulate the level of the signal as sampled by a reference microphone, which was the measurement microphone of the NTi Audio XL2. A standard STIPA measurement signal was played through the artificial head simulator. The head's mouth output level was tuned to ensure that the level corresponds to 70 dBZ at 1 m distance. The configuration of this measurement phase is depicted in **Figure 3.9**.

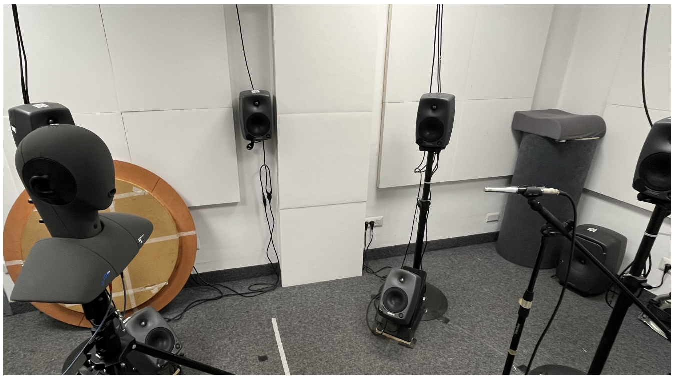


Figure 3.9: Calibration of the dummy head emissive mouth

Then, the head simulator has been recorded while reproducing the reference signal, using the NTi Audio XL2 calibrated microhpone, which was previously calibrated using the B&K 4231 calibrator.
The head-worn microphone's input gain was adjusted. By adjusting the virtual level of the head-worn microphone recordings to match the level recorded by the NTi microphone, we could assume that both microphones are calibrated to the same reference level. We also paid attention that the head-worn microphone could sample the performance of the singers without exceeding the limit of 0 dBFS, to avoid any clipping or microphone distortion. The distance from the artificial mouth to the microphone capsule was circa 4.5 cm (see **Figure 3.10**).

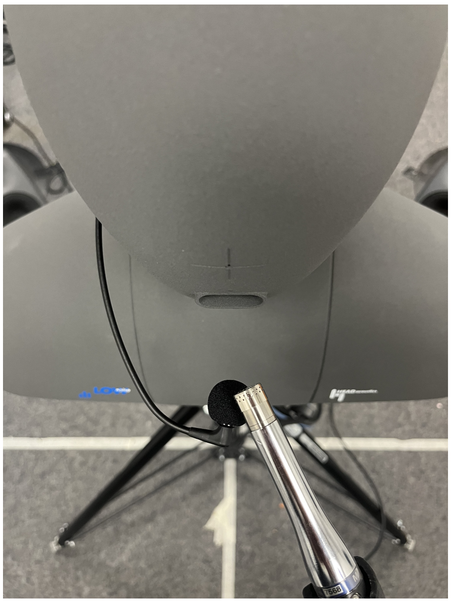


Figure 3.10: Positioning of the reference and head-worn microphones

The artificial head by itself is also capable of listening, thanks to a pair of microphones and human-like pinnaes and ear canal, that simulate the human hearing features. So, it could be used to simulate an experimental real-time auralization, as it would take place with a human user.
The artificial head was placed at the center of the loudspeaker array, as depicted in **Figure 3.11**. The same standard signal used before (STIPA standard signal, 70 dBZ @ 1m) was played through the artificial head's mouth simulator and recorded using the head-worn microphone, which was calibrated as described just before. Then, the recording was played through the auralization system, using the Bidule `player` node. By alternately convolving this signal with the modified IRs for the different churches, the reverberant tail of each environment was played through the loudspeaker array and reached the performers' ears. The output gain was then

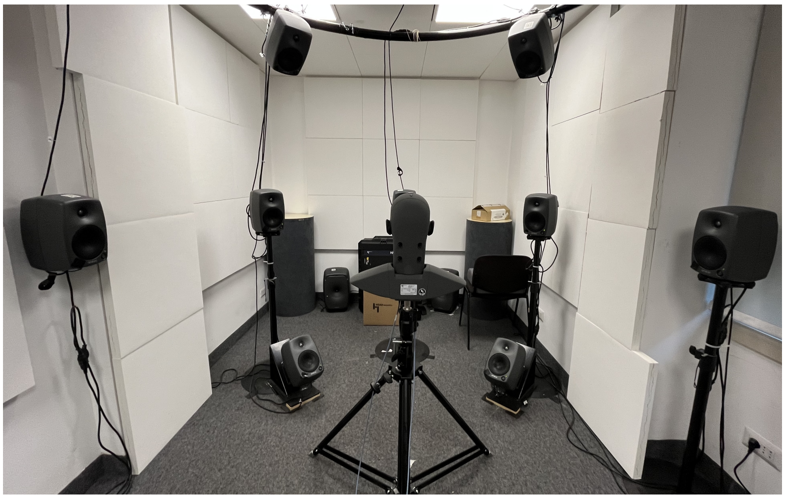


Figure 3.11: Dummy head simulator placed in the loudspeakers array

tuned to ensure that all the resulting sounds emitted from the loudspeaker array met a precise value, equal for each of them.

The level of the sound was measured using a built-in function of the ARTEMIS software, which can be connected to the artificial head to performe live monitoring and to record the input using the *HEAD labHSU* hub. The ear microphones were previously calibrated using the standard procedure described by the producer.

A minute-long file for each environment was recorded and analyzed using ARTEMIS and its built-in tools to calculate the A-weighted Equivalent Level. The output gain of the convolution system was then adjusted and the measuring procedure was repeated, until the resulting auralizations' equivalent mean level between the two ears matched the value of 65 ± 0.1 dBA. The results are shown in Table 3.1.

Environment	Left Ear [dBA]	Rigth Ear [dBA]	Mean Level [dBA]
Env 1	65.0	64.9	65.0
Env 2	64.9	65.0	64.9
Env 3	64.9	65.1	65.0
Env 4	65.0	65.2	65.1

Table 3.1: Equivalent A-weighted Levels for each auralization

3.3 Objective Parameters in Audio Space Lab

To characterize the auralized environments generated by the real-time convolution system, both omnidirectional and Oral-Binaural Room Impulse Responses (OBRIR) were acquired.

For the measurement of the omnidirectional IRs of the four auralized environments, a sweep signal (duration: 10 s, frequency range: 20 Hz–20,000 Hz, sample rate: 48,000 Hz) was convolved with the modified IRs of the churches, in which the direct sound component had been removed (see Section 3.1). The resulting sound was then reproduced through the loudspeaker array and recorded using an omnidirectional microphone positioned at the center of the array. This procedure was repeated three times for each of the four environments, allowing for the extraction of averaged room acoustics parameters in MATLAB, using the functions available in the `ita_toolbox()` suite. From the resulting omnidirectional IRs, RTs were obtained. The objective of these calculation was to obtain a set of RT values and compare them to the ones measured in the actual churches. The results are shown and discussed in the following chapters.

The OBRIRs were obtained by recording the same sweep signal, but played through the mouth of the artifial head simulator by HEAD Acoustics. The signal was reproduced through the mouth simulator of the dummy head, which had been previously calibrated to reproduce a Sound Pressure Level of 70 dBZ at a distance of 1 m. The resulting Sound Pressure Level of the emission of the sweep was also measured at 1 m of distance (using the NTi XL2 Audio meter), resulting in an equivalent level of 80.8 dBA at a distance of 1 m.

As for the calibration procedure detailed earlier, both the head simulator's emission and the real-time auralization were driven by a Bidule patch, while the recordings from the artificial head's ears were taken using Artemis. The ears had been calibrated and set to a maximum operative range of 122 dBSPL.

In **Figure 3.12** the measurement setup is depicted: the dummy head was placed in the sweet spot of the loudspeakers array, and equipped with the head-worn microphone. Three recordings for each auralized environment were taken, each including a tail of 7 seconds of silence to ensure the complete decay of the sound. After the recordings, the OBRIRs were obtained through a MATLAB script that performed a convolution with the inverse filter of the sweep, which was generated together with the sweep itself. The binaural parameters calculated after this procedure were the Room Gain (G_{RG}), the Vocal Support (ST_V) and the Mouth-Ear Decay Time ($\mathrm{DT}_{40,ME}$). These parameters were largely used by *Pelegrín-García et al.* and *Puglisi et al.* ([16, 17, 42, 43]), as they were found to be strongly correlated with the vocal comforte of talkers in different acoustics environments. In particular, they focussed on the analysis of vocal comfort of teachers in occupied and unoccupied classrooms, and the potentially bad effects that these rooms' acoustics could have

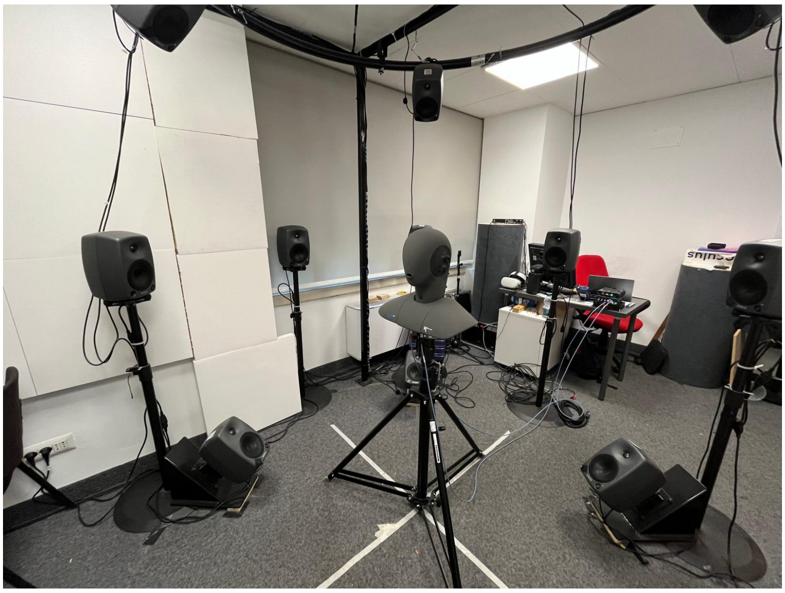


Figure 3.12: Dummy head simulator positioning for measuring procedures

on the wealth of the voice of the teachers.
In this work, the calculation of these parameters was performed in the frequency range covering the cotave bands centered on 125 Hz and 4000 Hz. This range was supposed by the author to be comprehensive of all the features of the human singing voice. Then, to obtain a single value comparable among the different environments, an arithmetical mean of the values in the octave bands centered on 500 Hz to 2000 Hz was computed.

Chapter 4

Subjective Investigation Protocol

In this chapter, the subjective investigation protocol adopted in the ASL is presented. A total of 18 participants took part in the study. The tests had a total duration of approximately 60 minutes, divided into three parts. In the final part, the participants' performances were recorded for subsequent analysis. The results of this analysis will be discussed in the following chapter.
The same procedure was repeated with the same soloist who performed in the actual churches. The results of the objective voice parameters and subjective responses will be analyzed in the following chapter.
The details of the testing procedures were refined through a preliminary study involving three volunteers. Subsequently, the final test protocol was confirmed and is presented in the following pages. A detailed description of the perceptual questionnaire, along with a discussion of the choice of descriptors used to guide participants in evaluating the audible features of the environments, is also provided.

4.1 Preliminary pilot study and refinements of testing procedure

The subjective investigation took place in the ASL during February 2025. Prior to this, three persons (two females and one male) participated in a preliminary test, which allowed for refinements in both the questionnaire and the testing procedure. Based on their feedback, the final testing procedure was structured as follows:

- A total duration of 40 minutes, divided into two smaller blocks of 20 minutes each;
- Three brief pauses, each lasting 5 minutes and placed between the blocks;
- The opportunity to freely explore the acoustic environments reproduced by the spherical loudspeaker array;
- The possibility to share comments and opinions about the experience without restrictions;
- The use of a questionnaire with fixed 5- or 7-point Likert scales to guide participants in evaluating their perception on the acoustic characteristics of the environments during the experiment;
- The recording of the performance of a well-known children's musical piece, chosen to be natural and effortless regardless of the participant's singing experience.

It is worth noting that all participants who took part in this subjective investigation were native Italian speakers or had multiple years of academic experience with the Italian language. Therefore, the questionnaires and instructions were provided in Italian.

4.2 Test Protocol

The total duration of 40 minutes of free exploration was chosen to ensure that each participant could experience and evaluate the acoustic characteristics of all environments, with an average of 10 minutes dedicated to each. During the pilot phase, it was observed that individuals with musical experience or a background in music or acoustics tended to require less time to complete the questionnaire. Nonetheless, the decision was made to allow ample time to minimize biases and prevent any pressure arising from time constraints.

Three breaks of five minutes were included to prevent both listening and vocal fatigue: in fact, a continuous vocal effort of 40 minutes could pose a risk to the vocal apparatus, particularly since no prior vocal warm-up was performed. This latter decision was made to ensure the participation of individuals without formal singing training while also avoiding the necessity of a trained vocal coach's constant presence. The inclusion of such additional elements would have introduced logistical challenges in organizing the tests and was deemed to be beyond the scope of this preliminary study.

Additionally, the breaks allowed participants to reset their auditory perception. Continuous exposure to artificial acoustic environments could have temporarily impaired their ability to reference the natural acoustic behavior of real-world spaces: while prolonged exposure might have enabled participants to become highly familiar with the artificial environments, it could have also diminished their ability to assess the realism of the reproduced acoustics in comparison to an actual physical space.

The participants were instructed to freely explore the acoustic environment in which they were placed. They were encouraged to vocally express themselves without hesitation, regardless of any potential self-consciousness that might limit them in using their full vocal range and volume. They were explicitly informed that the ASL was acoustically isolated from the rest of the building, ensuring that no one could hear their singing, except for the operator stationed at the PC controlling the reproduction system. Additionally, the operator was responsible for monitoring the input level of the head-worn microphone to prevent clipping, which could introduce distortions and unpredictable effects on the real-time auralization.

Also, the participants were instructed to freely compare the environments, by "moving" through them and comparing the perceived differences and similarities. To help this evaluation, they were told that they may pass from one environment to another simply by asking the operator and giving the id of the environment they wanted to move to. For statistical reasons, in fact, the order of the churches was randomized through different participants, so that the same environment id among different participants should not correspond to the same IR used for the real-time convolution. Nevertheless, the Bidule patch used for this real-time operation was designed to allow the operator to switch from one auralization to another in a span of just a couple of seconds.

It is important to note that the participants were instructed to keep their heads still for the entire duration of the experiment. This requirement was necessary because no head-tracking system was implemented, and any variation in head rotation could have caused a misalignment between the reproduction system and the listener's auditory perception, causing misleading and unpredictable effects on the realism of the auralization. Therefore, participants were asked to stimulate the acoustic environment only while facing forward, although they were free to move as needed when completing the questionnaire.

In the questionnaire, two Likert scales were used. At first, all questions adopted

a 5-points scale, but during the preliminary phase some of the participants complained about the lack of precision: in particular, when evaluationg the value of the acoustic descriptors (which will be presented in the following section), the 5-point scale was not enough to express the nuanced difference while directly comparing two environments. On the other hand, a 9-points Likert scale was considered by the designer of the questionnaire to be excessively detailed, leading to a potential loss of precision while evaluating the scores.
The final decision was to adopt a 7-points scale to rate the acoustic descriptors, while keeping a 5-points scale to evaluate the more general statements on Three-Dimensionality, Plausibility, Presence and Ease of Singing.

4.3 Anagraphic Data Collection

Before the start of the test, participants were asked to complete a brief questionnaire regarding their past experiences with immersive audio, acoustics, and musical knowledge. To maintain anonymity as much as possible, each participant was assigned a unique progressive ID. This ID was used solely for tracking the responses in both the questionnaire and the corresponding recordings.
From this initial questionnaire, infos about the past experience of the participants were asked, such as any previous knowledge about hearing or vocal tests, or whether they have past experience with the immersive audio. More, they were asked is they consider themselves as musicians, to rate their ability level as singers and/or as instrumentalist, and if they are part of any choir, band or orchestra, their repertoire. In the end, they were asked whether they have any experience of singing for a public or in an environment with long reverberation times. A Likert scale with 7 points was used to rate their knowledge about acoustics, music, and singing.

4.4 Questionnaire

The questionnaire was designed to guide the participants during the evaluation of the characteristics of the acoustic environment in which their voice were auralized. It contained a total of 5 questions for each of the four environments. The questionnaire used during the experimental protocol can be found in **Appendix A**, in its original Italian version.
The first four questions focussed on the sense of Three-Dimensionality, Realism, Presence and Ease of Singing. For each of the statements, the participants were asked to rate their agreeement on a 5-point Likert scale, ranging from "Absolutely disagree" to "Absolutely agree".

The last question presented seven descriptors linked to the acoustic characteristics of the environment and the way the sound was perceived by the participants. For each of the descriptors, the participants were asked to rate their agreement on a 7-point Likert scale. The ranges vary based on the meaning of the descriptor, but were sorted to present a badder sense at the left, while getting a positive meaning the more the scale proceeds towards the right. To have a more precise localization of the values, a mid, neutral value was also inserted.
The descriptors were:

- **1. Colorfullness**: the alteration of the sound timbre due to filtering effects, resonances and distrosions of the reproduction system or introduced by the environment itself;

  - **Dark**: the sound is like reproduced by a disco speaker, which emphasizes the low tones;
  - **Mid**: normal reproduction *(intended as the expected reproduction from the real environment, Translator's note)*;
  - **Bright**: the sound is like reproduced by a low-quality speaker, which emphasizes the high tones.

- **2. Spatiality**: the perception of the geometrical dimensions of the acoustic venue which sourrand the listener;

  - **Small**: a small room;
  - **Medium**: a medium room;
  - **Big**: a very big room.

- **3. Reverberation**: the extension of the sound caused by the reflections of the sound on the surfaces of the environment, which influence the perceived dimensions and depth of the acoustic environment;

  - **Dry**: a recording studio *intended as a treated room, with low reverberation, Translator's note*;
  - **Mid**: normal room;
  - **Reverberant**: cathedral.

- **4. Clarity**: the possibility of clearly distinguish notes played in a quick succession, which is influenced by the presence of background noise, distorsion or lack of quality in the audio source;

  - **Messy**: the sound are not clearly distinguishable one another, causing a general confusion/noise;
  - **Mid**: the sounds are intelligible but there's still a distrubance of some sort;
  - **Clear**: the sounds are intelligible and no distrubance is present.

- **5. Pitch**: the perceived height of the sound, determined by its fundamental frequency;

  - **Deep**: the notes are generally deeper;
  - **Mid**: the notes are the same as they have been emitted;
  - **High**: the notes seam to be generally higher.

- **6. Vibrato**: a periodic modulation of frequency of the sound, often used as an expressive tool in singing and playing;

  - **Hard**: the environment makes it harder to vibrate;
  - **Mid**: the environment does not influence your vocal vibration;
  - **Easy**: the environment makes it easier to vibrate.

- **7. Velocity**: the speed of the sound or the sequence of sounds is reproduced, which influences the perceived duration or rithm of the sound;

  - **Slower**: it seams to you that you slowed down your singing;
  - **Mid**: you did not notice any change in the speed of your singing;
  - **Faster**: it seams to you that you accelerated your singing.

It is worth noting that these acoustic descriptors were presented in Italian, and here is presented the literal translation of the captions provided, which may lead to cultural misunderstandings and imperfect adaptations towards other languages.

For each environment, a section was dedicated to free comments. The participants were encouraged to express any opinion they had about their feelings or emotions while perfoming in each auralized environment, and their comments were collected both using the free comment section on the questionnaire and by the operator, who was present in the room during the entire duration of the test. No replying comments was ever made by the operator, to avoid bias on the free expression of the performer.

The majority of the participants chose to freely comment by voice the acoustic impression they were experiencing, rather than writing them in the booklet. A fewer number of them, though, preferred to write them, whether it was for a more precise and structured expression of their feelings, or for a more private and intimate way of sharing their opinions.

On the last page of the questionnaire, a section was dedicated to collect impressions about similarities and differences among the four environments. The participants were asked to evaluate the overall difference among all the acoustics they experienced during the test, on a 5-point Likert scale ranging from “Absolutely similar” to “Absolutely different”. Then they were asked to describe the main differences through the same 7 descriptors used in the previous questions, with an additional “Others” options in which they could express any other mismatching characteristic which was not listed. The same form was used to collect opinions about the similarities among the environments.

At the end, they were asked to rate their favourite environments, and to express the reasons of their choice. They were instructed to focus on the environments in which they could sing in the easiest and most comfortable way, but any other guidance parameter for the choice was not provided, leaving the participant to evaluate freely the characteristics that they thinks would contribute the most on their delivery in each of the auralized acustics they experienced.

In the last block of every test, the participants were recorded while singing a well-know, child song: "Happy Birthday to You" (sang in English) was chosen for the simple melody, which is analogous to the Italian version. Every participant granted that they knew well the text of the English version. Every performance was recorded with the head-worn microphone, and trimmed to ensure that nothing but the singing part was memorized. These recordings were analyzed by means of MATLAB scripts, as it will be discusses in the following chapter.

Chapter 5

Results and Discussion

In this chapter, all the datas collected during the various phases of this thesis are reported and analyzed. Starting from the room acoustics parameters sampled in the four churches, the chapter presents the sets of measurements obtained from the auralization in the ASL. Then, voice parameters from the recordings of the soloist in the actual environments and the ones obtained from the recordings performed during the subjective investigation are presented.
Lastly, a correlation has been made considering the voice parameters as dependent variables of the acoustical parameters, and presented as a series of plots in the final part of this chapter.

5.1 Objective Acoustical Parameters in Churches

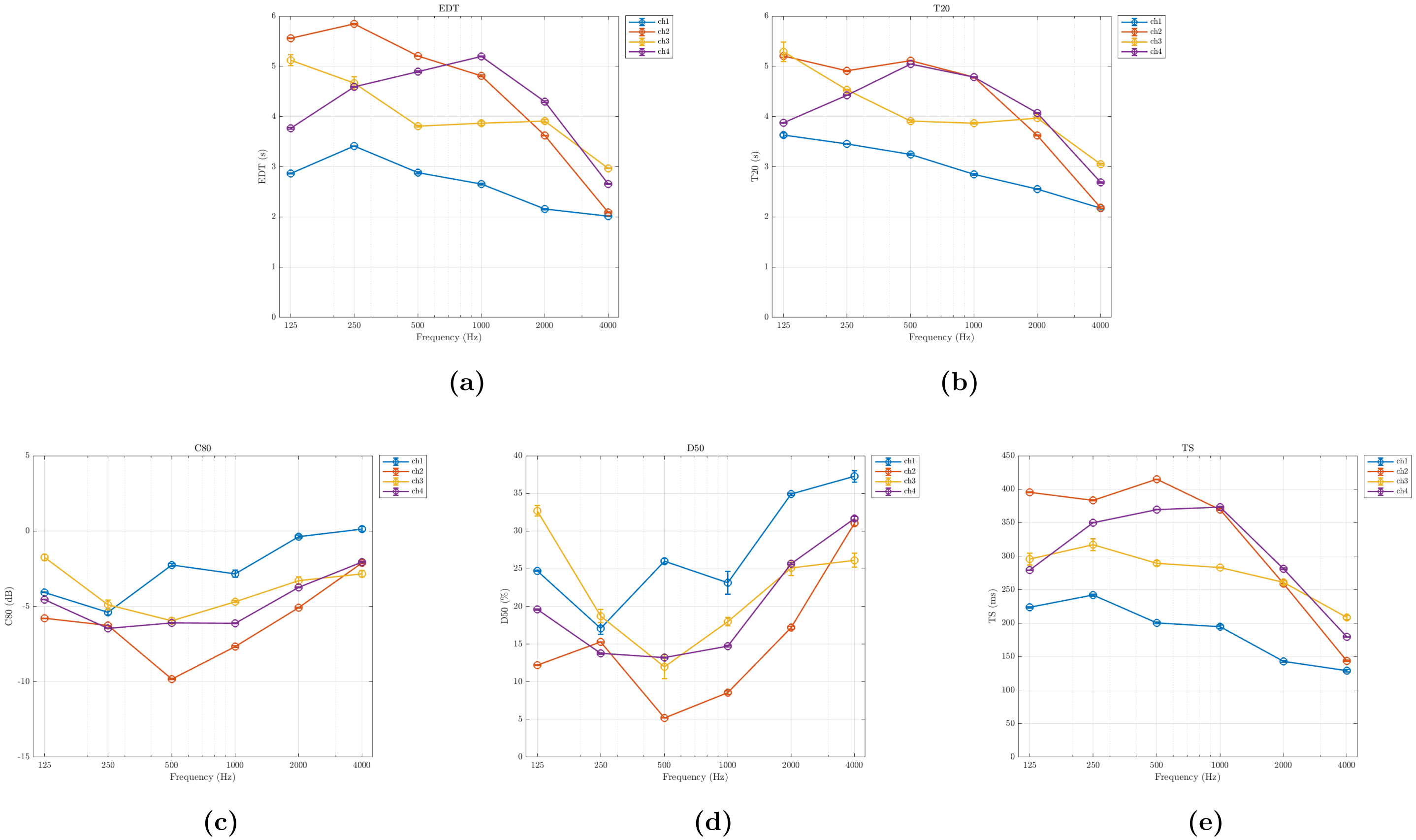
Figure 5.1 depicts the values of some Room Acoustics Parameters obtained in the reference position in each of the four churches. Three iterations of the measuring process using a sweep signal were performed, and mean values for each octave bands have been computed. Parameters are presented in the octave bands 125 Hz to 4000 Hz. Single values (reported in Table 5.1) were derived performing an arithmetical mean on the octave bands 500 Hz and 1000 Hz.

BR and TR were calculated using Formulae 2.4 and 2.5. Values were obtained from the three EDTs calculated with the three iterations, and then the mean values and standard deviation were computed.

IACC, IACC$_{early}$ and IACC$_{late}$ values were obtained from the 3OA IRs in the reference positions, in octave bands. Single values were obtained by averaging the values in octave bands from 125 Hz to 4000 Hz.

The four churches present significant differences in the majority of these parameters. In particular, Church 1 present the less reverberant acoustics, which also results in greater C_{80} and D_{50} values. On the other hand, Church 2 and Church 4 present similar EDT and T_{20}, except for the lower bands where Church 4 shows reduced reverberation times. Church 3 has an intermediate behaviour, being more reverberant at higher and lower frequencies and drier around 500-1000 Hz.

IACC values present lower differences between the four churches, which are lower than the JND of 0.075 reported in [31]. For this reason, these parameters will not be considered in these analysis.



(a)

(b)

(c)

(d)

(e)

Figure 5.1: Room Acoustics Parameters in the churches

Parameter	**Church 1**	**Church 2**	**Church 3**	**Church 4**
$\mathrm{EDT}_{500-1000\mathrm{Hz}}$ [s]	2.77 (0.01)	5.00 (0.01)	3.84 (0.01)	5.05 (0.01)
$\mathrm{T20}_{500-1000\mathrm{Hz}}$ [s]	3.05 ($<$ 0.005)	4.95 (0.01)	3.89 (0.02)	4.91 ($<$ 0.005)
$\mathrm{C80}_{500-1000\mathrm{Hz}}$ [-]	-2.54(0.13)	-8.74 (0.04)	-5.31 (0.12)	-6.10 (0.02)
$\mathrm{D50}_{500-1000\mathrm{Hz}}$ [%]	24.58 (0.91)	6.87 (0.11)	14.98 (1.08)	13.97 (0.07)
$\mathrm{TS}_{500-1000\mathrm{Hz}}$ [ms]	197.45 (1.86)	392.26 (0.88)	286.37 (2.12)	371.69 (0.53)
BR [-]	1.13 (0.01)	1.14 ($<$ 0.005)	1.28 ($<$ 0.005)	0.83 ($<$ 0.005)
TR [-]	0.75 (0.01)	0.57 ($<$ 0.005)	0.90 ($<$ 0.005)	0.69 ($<$ 0.005)
$\mathrm{IACC}_{fulltime,0.125-4\mathrm{kHz}}$ [-]	0.54 (0.01)	0.50 (0.01)	0.49 (0.01)	0.50 ($<$ 0.0050.03)
$\mathrm{IACC}_{early,0.125-4\mathrm{kHz}}$ [-]	0.66 (0.01)	0.66 (0.01)	0.57 (0.02)	0.65 (0.03)
$\mathrm{IACC}_{late,0.125-4\mathrm{kHz}}$ [-]	0.47 ($<$ 0.005)	0.46 (0.01)	0.46 (0.02)	0.47 (0.01)

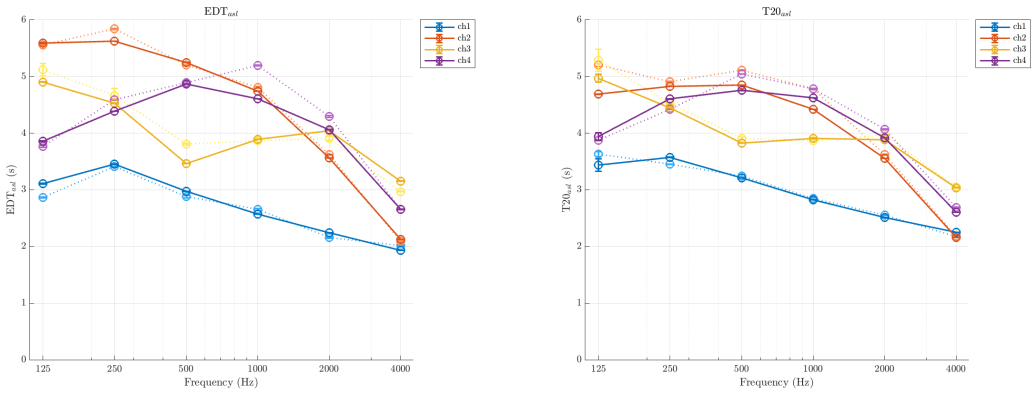
Table 5.1: Room acoustics parameters: single values, *mean (SD)*

5.2 Objective Acoustical Parameters in ASL

5.2.1 Reverberation Time in ASL

Both EDT and T_{20} were calculated from the IR measured in the ASL. These results can been compared to the measured values obtained in the real churches, as presented in **Figure 5.2**.

In this plot, the values obtained in the ASL *(straight lines)* and the values obtained in the actual environments *(dotted lines)* present evident differences, particularly for what concerns EDT in Church 4 and T_{20} in Church 2, while the other curves present a more similar behaviour in the two series. Church 1 among all the others present pretty similar behaviours in both the actual and the virtual environment. Single values were obtained averaging the value in the octave bands from 500 Hz to 1000 Hz, and presented in Table 5.2. The difference between the values in the real and the auralized environments have been computed and compared to a reference threhold calculated as the 5% of the single value in the real churches. The values exceeding this threshold are presented in a bold font, and correspond to T_{20} in Church 2 and EDT in Church 4: in both cases, the values measured in the ASL are greater than the ones measured in the correspondent church. All the other values are lower than this threshold.



(a) **(b)**

Figure 5.2: Auralized RT in the ASL *(continuous line)* compared to the ones measured in the churches *(dotted line)*

Envs	**EDT** [s]	**T20** [s]	**EDT$_{asl}$ - EDT$_{ch}$**	**T$_{20,asl}$ - T$_{20,ch}$**
Env. 1	2.78 ($<$ 0.005)	3.02 ($<$ 0.005)	-0.01	0.03
Env. 2	4.99 ($<$ 0.005)	4.64 ($<$ 0.005)	0.01	**0.31**
Env. 3	3.68 ($<$ 0.005)	3.87 ($<$ 0.005)	0.16	0.02
Env. 4	4.74 ($<$ 0.005)	4.69 ($<$ 0.005)	**0.31**	0.22

Table 5.2: Auralized Reverberation Times: single values, *mean (SD)*

5.2.2 Binaural Parameters

Binaural parameters were recorded in the ASL using a dummy head simulator. The results are presented in Table 5.3. It can be observed that neither G_{RG} nor ST$_V$ show significant differences across the various environments and will therefore not be considered in the following sections. Conversely, DT$_{40,ME}$ exhibits noticeable differences among the four environments, which align with the expected ranking of reverberation as indicated by the EDT and T_{20} values.
These finding were interpreted by the author as an indicator that the four auralized environments supply a similar support at the performers' ears, reguardless of the differences in the length of the reverberant tails, which are maintained as indicated by the different values of DT$_{40,ME}$.

Parameter	**Church 1**	**Church 2**	**Church 3**	**Church 4**
G$_{RG}$ [dB] - SX - mean (SD)	0.40 (< 0.005)	0.40 (< 0.005)	0.40 (< 0.005)	0.40 (< 0.005)
G$_{RG}$ [dB] - DX - mean (SD)	0.35 (< 0.005)	0.35 (< 0.005)	0.35 (< 0.005)	0.35 (< 0.005)
G$_{RG}$ [dB] - BN - mean (SD)	0.38 (< 0.005)	0.38 (< 0.005)	0.38 (< 0.005)	0.38 (< 0.005)
ST$_v$ [dB] - SX - mean (SD)	-10.79 (< 0.005)	-10.79 (< 0.005)	-10.79 (< 0.005)	-10.79 (< 0.005)
ST$_v$ [dB] - DX - mean (SD)	-11.35 (< 0.005)	-11.35 (< 0.005)	-11.35 (< 0.005)	-11.35 (< 0.005)
ST$_v$ [dB] - BN - mean (SD)	-11.07 (< 0.005)	-11.07 (< 0.005)	-11.07 (< 0.005)	-11.07 (< 0.005)
DT$_{40,ME}$ [s] - SX - mean (SD)	0.53 (< 0.005)	0.95 (0.01)	0.73 (< 0.005)	1.01 (0.01)
DT$_{40,ME}$ [s] - DX - mean (SD)	0.47 (< 0.005)	0.85 (< 0.005)	0.65 (< 0.005)	0.90 (< 0.005)
DT$_{40,ME}$ [s] - BN - mean (SD)	0.50 (< 0.005)	0.90 (0.01)	0.69 (< 0.005)	0.95 (< 0.005)

Table 5.3: Binaural acoustical parameters: single values, *mean (SD)*

5.3 Professional Singers in churches

5.3.1 Soloist Voice Parameters

The recordings of the soloist performing the excerpt from “Crucifige” were analyzed, and the results are presented in the following pages. The recordings were performed in the four churches (listed as *Ch1* to *Ch4*) and in the anechoic chamber of Politecnico di Torino (labeled as *Ach*).

Due to the setup of the recordings, considerations regarding the SPL of the singer during the performance were not evaluable, along with considerations on the TD and the varying lengths of pauses.

Figure 5.3 depicts the distribution of F_0 values across the four churches and the anechoic chamber. The peak value indicates the most recurrent note, which may vary depending on the singer’s ability to maintain intonation during the execution.

Figure 5.4 shows the difference between the F_0 values and the reference note *FA3*—corresponding to a frequency value of 349 Hz—chosen as the note that best represents the tonality of the performance. Churches 1 and 3 displayed lower values, as the F_0 of the recordings in these churches tended to deviate further from the reference note. Additionally, the standard deviation of the F_0 values in these two environments was greater than that observed for Churches 2 and 4, indicating a larger range of values over time.

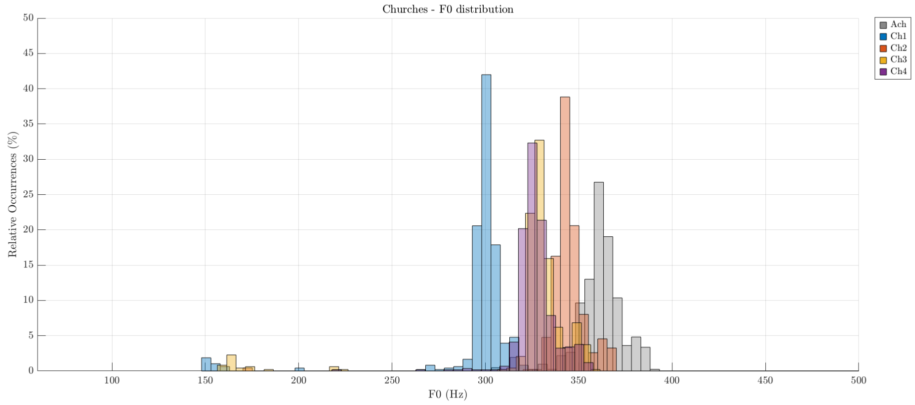


Figure 5.3: Distribution of F0 - Soloist across Churches

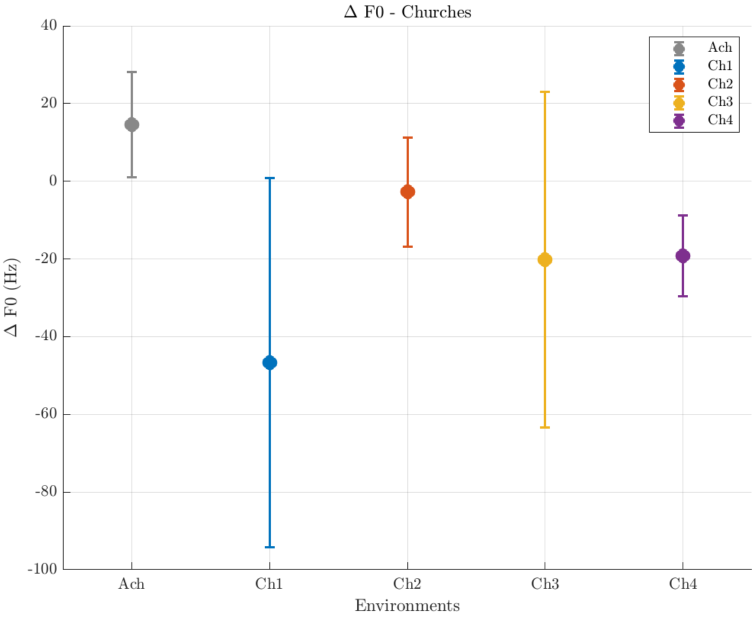
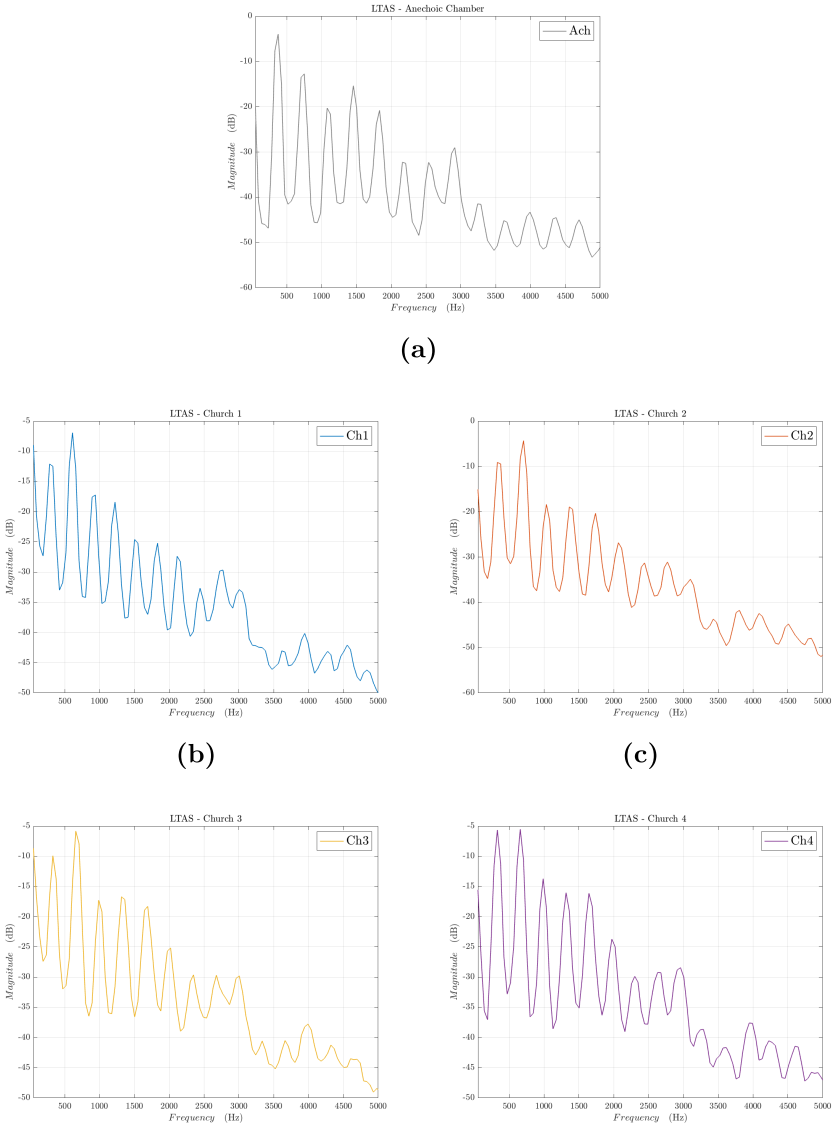


Figure 5.4: Variation of F0 in Soloist across churches: mean and standard deviation

The LTAS for each of the recordings were computed to verify the presence of the Singers' Formant, which is a prominence in the signal spectral envelope in the range from 2000 Hz to 4000 Hz. As reported in **Figure 5.5**, no clear prominence can be observed.



(a)

(b)

(c)

(d)

(e)

Figure 5.5: LTAS of the Soloist in the four churches

Some vocal features were extracted from the LTAS of the recordings.
First of all, the SPR was computed as in Formula 2.24. The mean values and standard deviations are reported in **Figure 5.6**. Values do not show clear differences among the different churches. Also, values for FCP, ΔB$_{low}$ and ΔB$_{high}$ were computed. These values show more significant differences among the four environments.

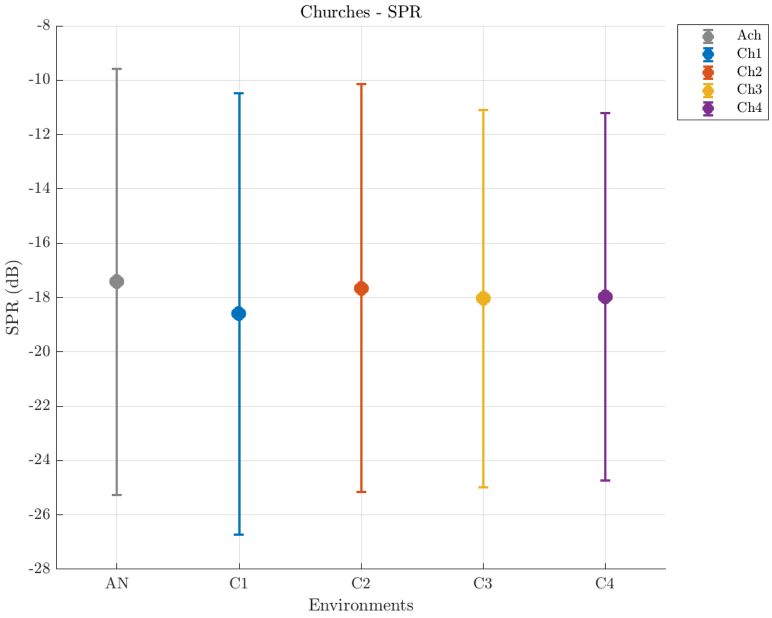


Figure 5.6: SPR of Soloist across churches: mean and standard deviation

Parameter	Ach	Ch 1	Ch 2	Ch 3	Ch 4
FCP [dB]	7.4	6.6	4.9	5.5	6.1
ΔB_{low} [dB]	-9.5	-11.9	-10.4	-8.5	-7.4
ΔB_{high} [dB]	-29.7	-28.4	-30.9	-27.9	-27.5

Table 5.4: Soloist voice parameters: LTAS features

CPPS was also computed. The distribution of values obtained from the recordings in the four churches is shown in **Figure 5.7**, while mean values and standard deviations are reported in Figure 5.8. Both graphs indicate variations in CPPS across the five recordings, with the highest mean value observed in Church 4 and the lowest in Church 1.

The mean value across all four environments is approximately 12 dB. However, it is important to note that assessments on the vocal health status based on CPPS yield more reliable results when applied to sustained spoken vowel recordings. Consequently, the values presented here may not necessarily correlate with the singer's vocal health.

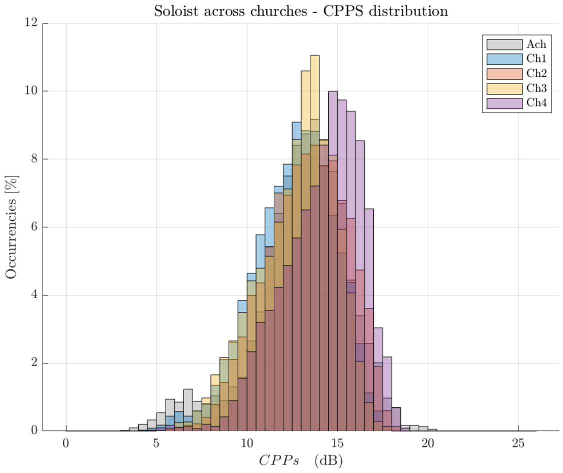


Figure 5.7: Distribution of CPPS values of Soloist across churches

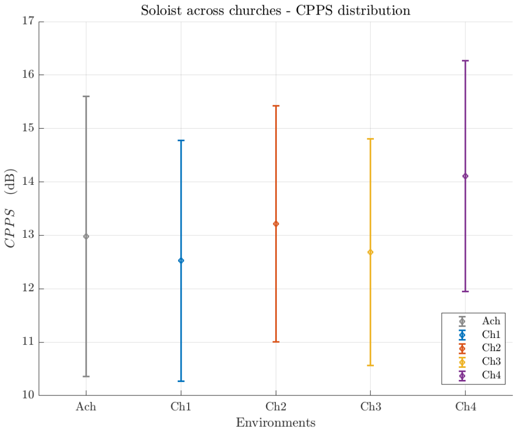


Figure 5.8: CPPS of Soloist across churches: median and standard deviation

5.3.2 Free Comments on Actual Churches

All the professional singers gave a series of free comments on the experience in each of the four churches. They shared the ease of singing in each of them, and created a preference ranking for the environment in which they liked the most to perform.

- Church 1 - **GRADE:** –:
  - very difficult space
  - dispersive space
  - pianissimo is optimized
  - need to find a relation between singers
  - need to prolong the singing
- Church 2 - **GRADE:** ++:
  - enriches the harmonics
  - need to create longer silences
  - less intelligible for the speech included in one piece of the performance
  - need to lower the sound level of the pronounced words
  - need to focus the sound
- Church 3 - **GRADE: -**:
  - cuts-out the mid frequency harmonics, which should be compensated with the singers' voices
  - compresses and dries the high pitch harmonics
  - need to pronounce the "a" sound with more "o" sound to soften the overall output
- Church 4 - **GRADE:** ++:
  - enriches the harmonics with the risk to overwhelm the principal sound
  - rich with its own sounds and voices according to the position of the singer
  - need to keep longer the end of the words

Singers reported a preference for Church 2 and Church 4, which, despite their differing characteristics, were rated as equally advantageous for performance. Conversely, Church 1 received the lowest rating, as it was perceived as overly dispersive, making it difficult for the singers to hear each other. Church 3 required the singers to compensate for a perceived lack of mid-frequency harmonics and a detrimental effect on higher-pitch harmonics, necessitating additional vocal effort. For this reason, they rated it as overall uncomfortable.

5.4 Subjective Investigation in ASL

In this section, data gathered during the subjective investigation in the ASL are presented. The voice parameters were obtained by analyzing the recordings from the third block of the experimental protocol, in which the amateur participants were asked to sing "Happy Birthday to You" in English once in each of the four auralized environments (labeled as *ch1* to *ch4*). Additionally, a recording of the song performed in the ASL with the auralization engine turned off was collected (labelled as *asl*). Due to recording issues, the recordings in this last configuration for participants 1 and 2 had to be discarded and were not considered in the analyses presented in the following sections.
In the final part of this section, answers to the questionnaire are also presented. As already explained in the proper chapter (see Chapter 4), the questionnaire was filled by the participants during the first two blocks of the experimental protocol, and was used to collect impressions of the perceptual characteristics of each of the four auralized environments.

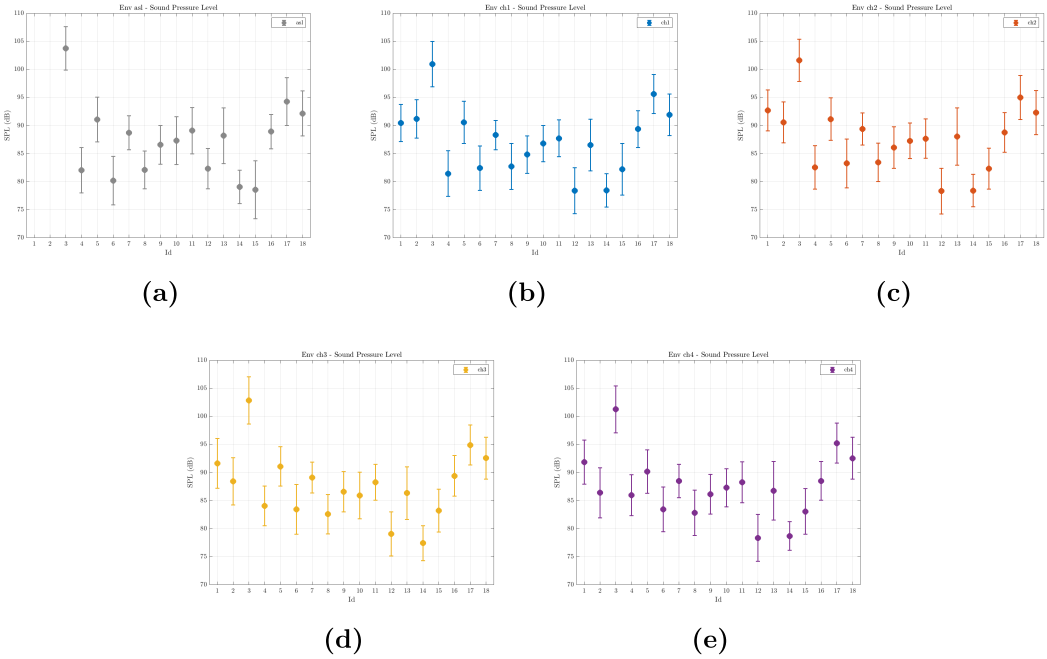
Some generic informations about the paticipants are presented in Tabel 5.5. The sample was compesed by 18 people: 14 Males, 3 Females, 1 Non-binary. Their age ranged from 20 to 63 years old (mean: 26.9, SD: 9.7) The majority of them (11/18) already had experience with Spatial Audio, either for academic use or personal knowledge. They were all native Italian speakers, except Volunteer 15 who is Spanish native speaker but has a multi-year experience with the Italian language.
The majority of them had previous experience in soloist or choral singing, but had never taken actual singing classes.

Id	Gender	Age	Voice Register	Exp. Spatial Audio
1	M	25	Baritone	Y
2	M	24	Baritone	Y
3	M	33	Baritone	Y
4	M	20	Tenor	N
5	M	24	Bass	Y
6	M	20	Tenor	N
7	F	24	Alto	N
8	F	27	Mezzo	Y
9	N.B.	24	Alto	Y
10	M	23	Bass	N
11	M	23	Tenor	N
12	M	63	Baritone	N
13	M	29	Baritone	Y
14	M	21	Baritone	N
15	M	33	Bass	Y
16	M	24	Tenor	Y
17	F	23	Alto	Y
18	M	24	Baritone	Y

Table 5.5: Informations about the volunteers for the subjective investigation

5.4.1 Voice Parameters from Tests

The SPL measurements are reported in **Figure 5.9**, while a mean between all the tests for the same environment is presented in **Figure 5.10**. A slight difference is observable in the mean value across the environments.



(a) **(b)** **(c)**

(d) **(e)**

Figure 5.9: Subjective Investigation: SPL acorss the enviroments (mean and standard deviation)

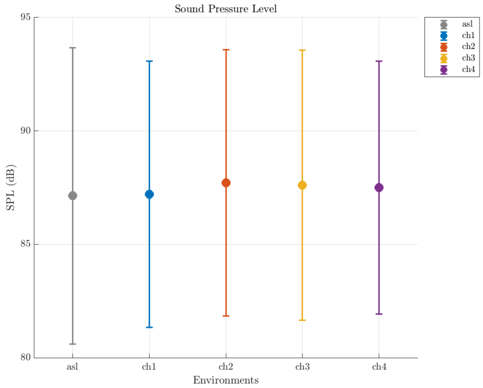
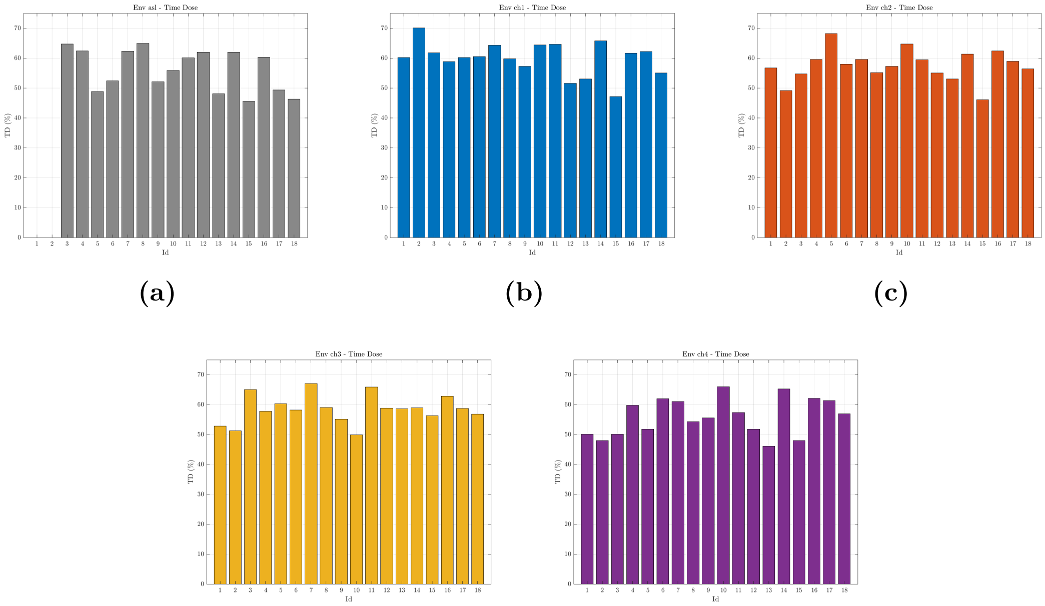


Figure 5.10: Subjective Investigation: SPL (mean and standard deviation)

Results for the TD are presented in **Figure 5.11**, while means between the tests for each environment are depicetd in **Figure 5.12**. From the latter, a clear difference is observable: TD is higher in enviroments *ch1* and *ch3*, while it is lower in *asl* and *ch4*. *ch2* has an intermediate value.



(a) **(b)** **(c)**

(d) **(e)**

Figure 5.11: Subjective Investigation: TD across environments

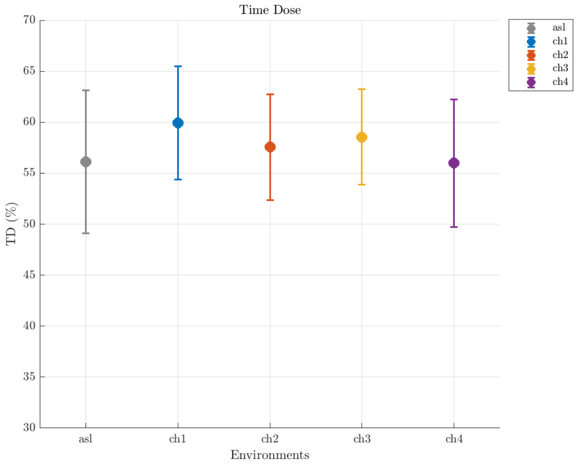
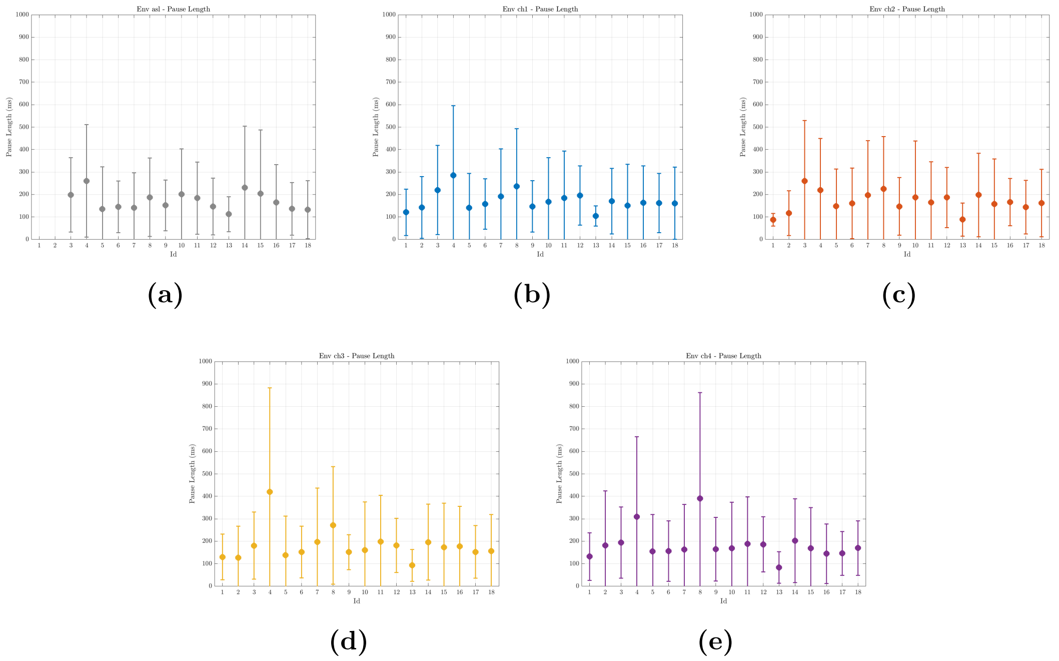


Figure 5.12: Subjective Investigation: TD (mean and standard deviation)

Results on the length of the pauses in the different environments are reported in the following **Figures 5.13** and **5.14**. On an average between all the participants in the five environments, *ch3* nd *ch4* present the longer pause times. Also, these two environments present the larger standard deviations, which may be read as an index of greater variation of these parameters.



(a) **(b)** **(c)**

(d) **(e)**

Figure 5.13: Subjective Investigation: Pause Length (mean and standard deviation)

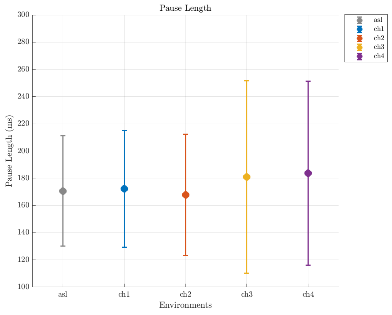
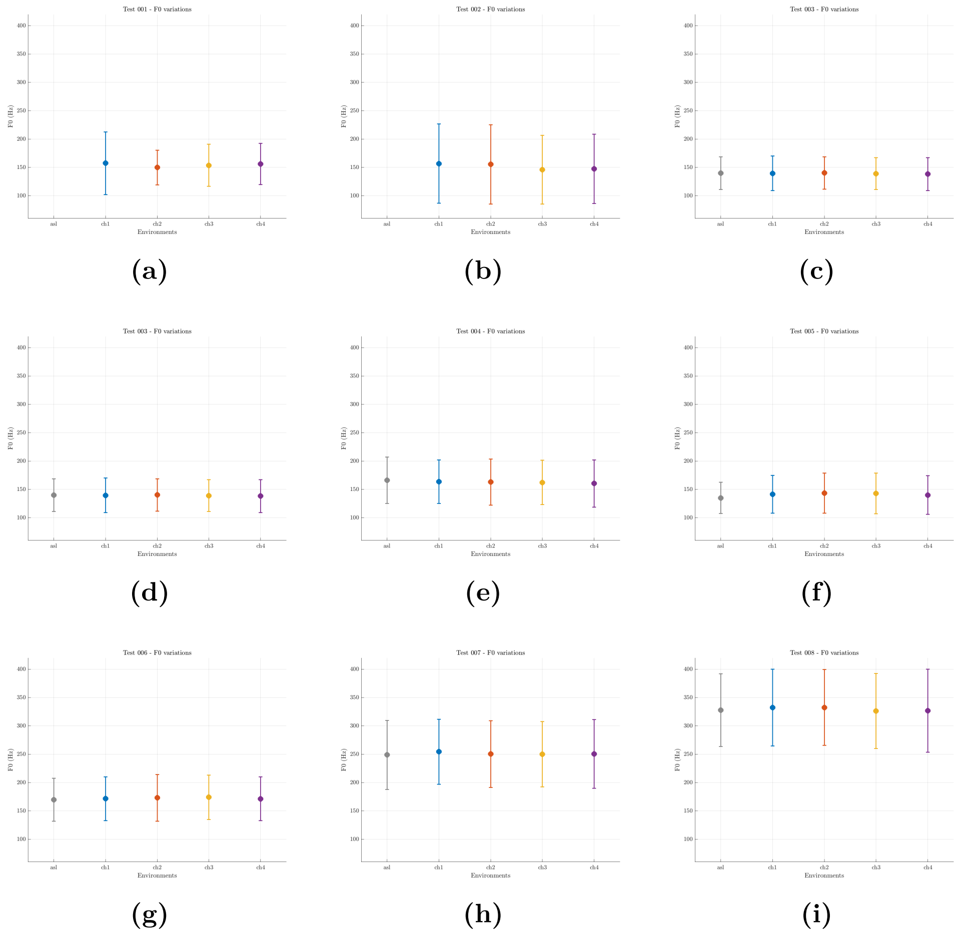


Figure 5.14: Variation of Pause Length across environments: mean and standard deviation

The F_0 of the singing excerpts was analyzed, and the values for each singer are presented in **Figure 5.15**. It can be observed that participants tent to maintain the reference note regardless of the influence of the auralized environment in which they are performing. The noticeable differences among the plots of different volunteers can be attributed to the different reference notes assigned to each participant, which were chosen based on their vocal register for optimal comfort.
Figure 5.16 shows the mean F_0 variations relative to the reference note given to each singer. The behaviour is consistent across all environments, with maximum differences of only a few Hz.



(a)

(b)

(c)

(d)

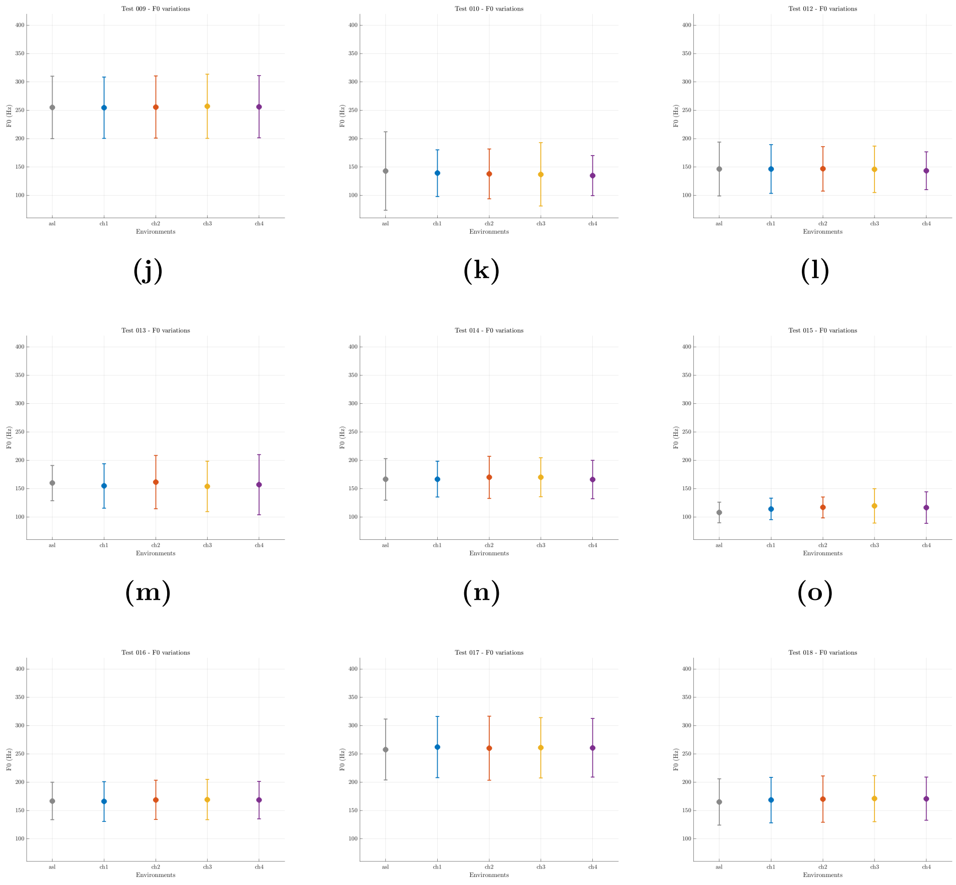
(e)

(f)

(g)

(h)

(i)



(j) (k) (l)

(m) (n) (o)

(p) (q) (r)

Figure 5.15: Subjective Investigation: F0 distribution

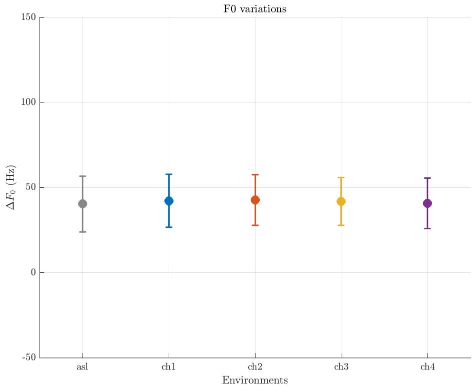
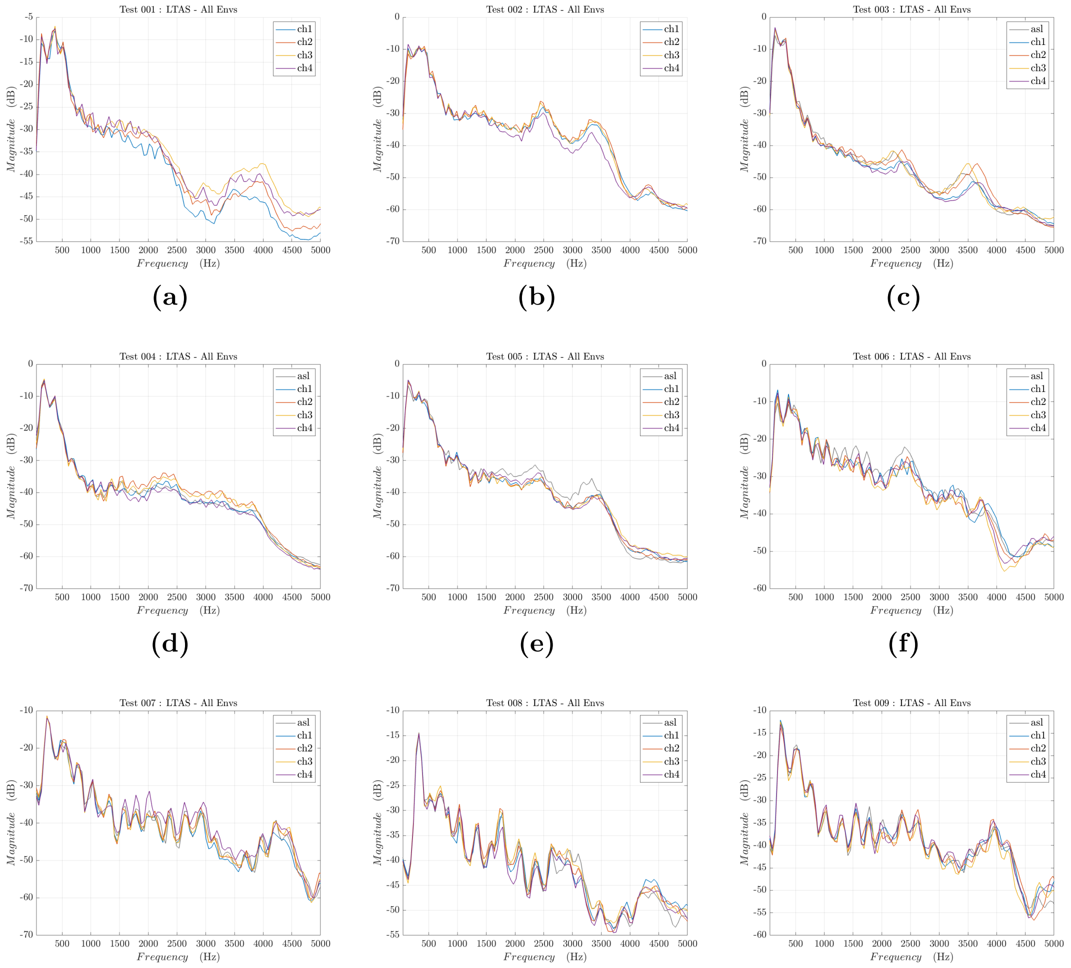


Figure 5.16: Variation of F0 to the given note: mean and standard deviation

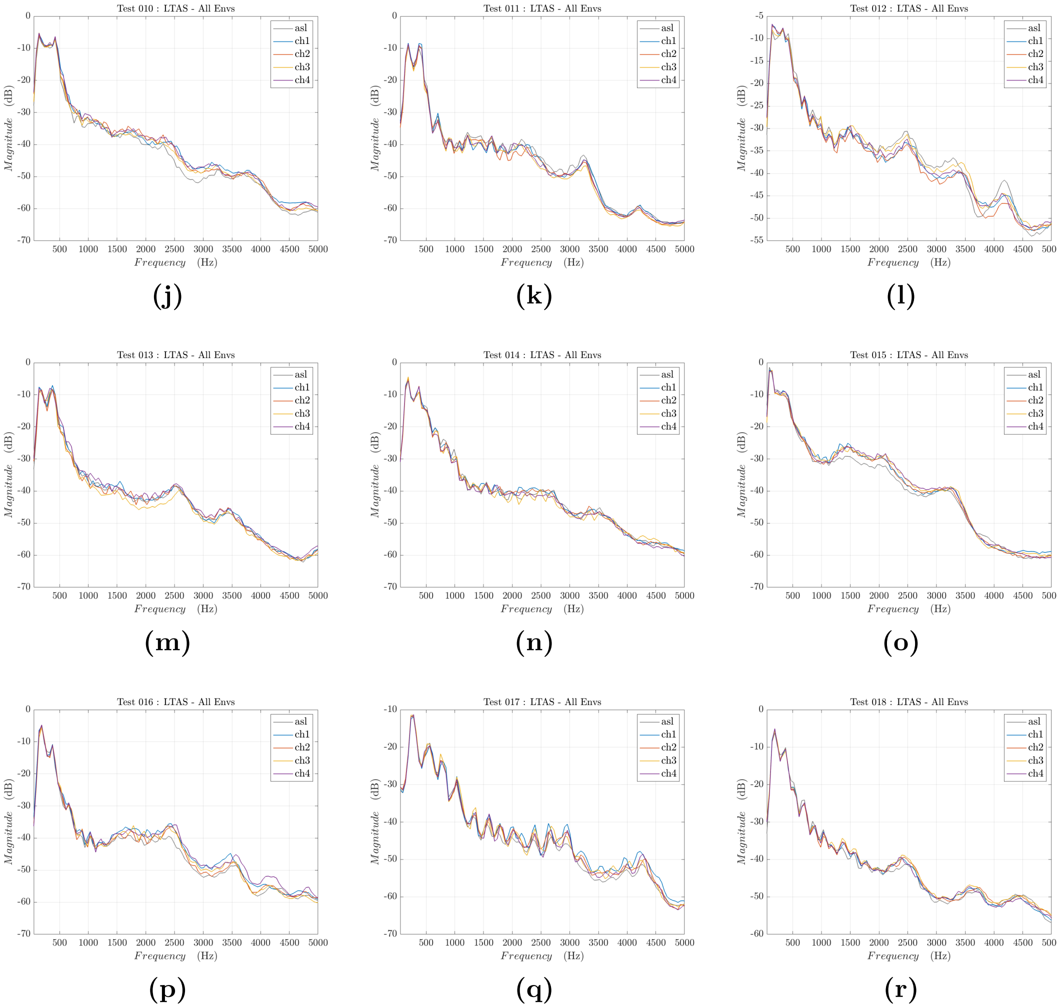
Figure 5.17 presents the LTAS graphs for each volunteer. In most cases, the plots are comparable across the five environments, although some subjects exhibit noticeable differences, particularly in the frequency range from 1500 Hz to 4000 Hz. In the majority of cases, a single, clear formant prominence in the 2000 Hz to 4000 Hz range could not be identified. The singer's formant, which typically appears within this range, was not clearly observed in these amateur singers.



(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g) **(h)** **(i)**



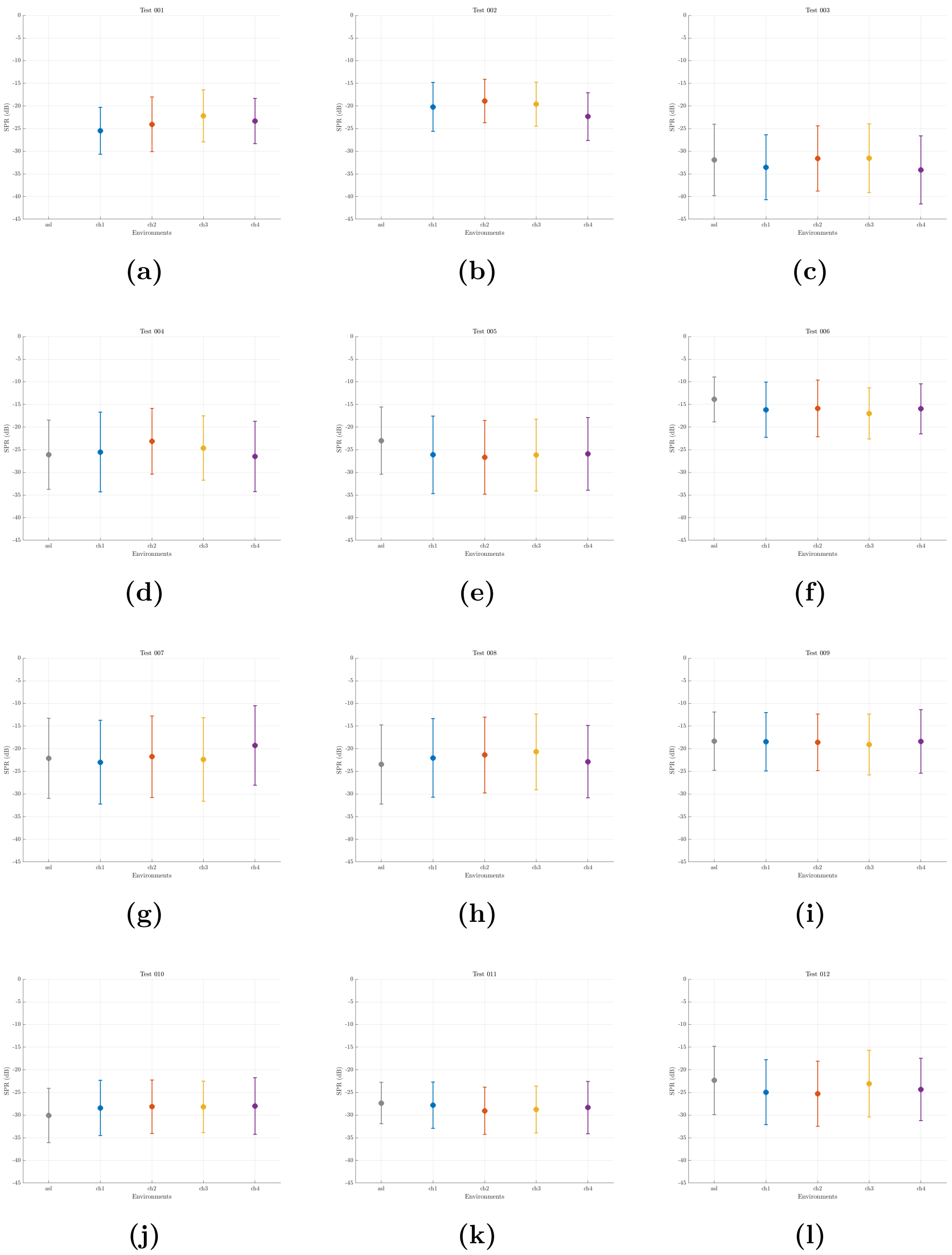
(j) **(k)** **(l)**

(m) **(n)** **(o)**

(p) **(q)** **(r)**

Figure 5.17: Subjective Investigation: LTAS

SPR has been computed as an index of ringing quality in the singing voice. **Figure 5.18** depicts differences for each amateur singer among different environments, while **Figure 5.19** presents mean values among the volunteers in the five environments. It can be clearly seen that the mean value is pretty similar between the different environments.

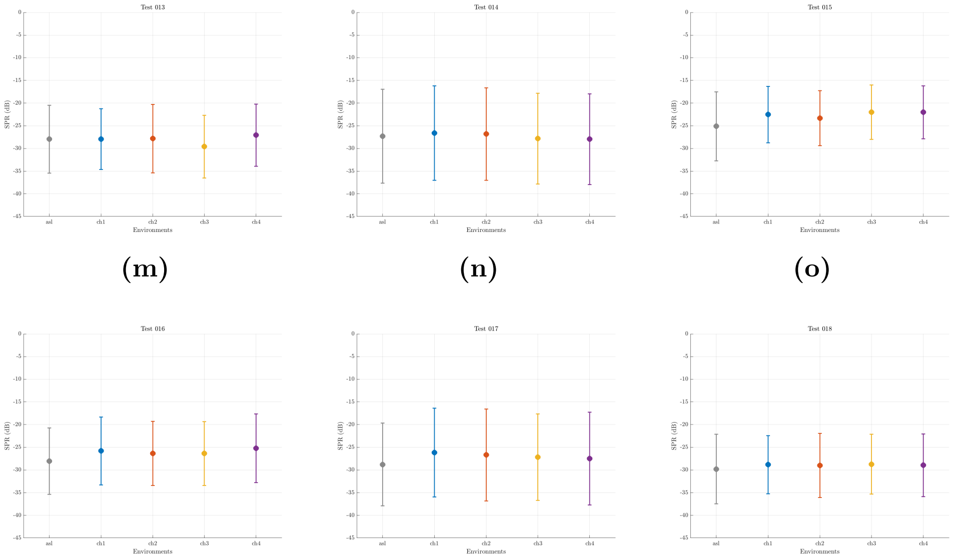


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)



(m) **(n)** **(o)**

(p) **(q)** **(r)**

Figure 5.18: Subjective Investigation: SPR

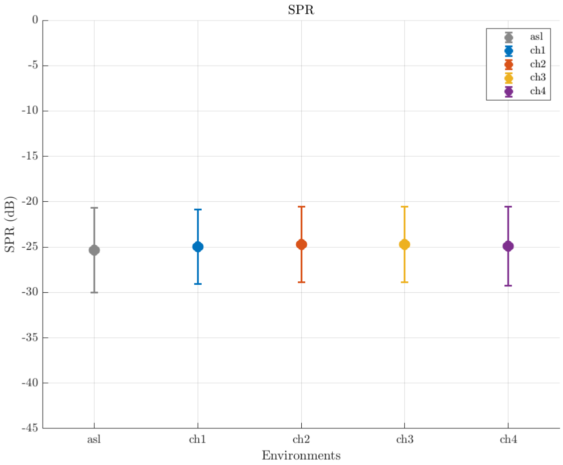
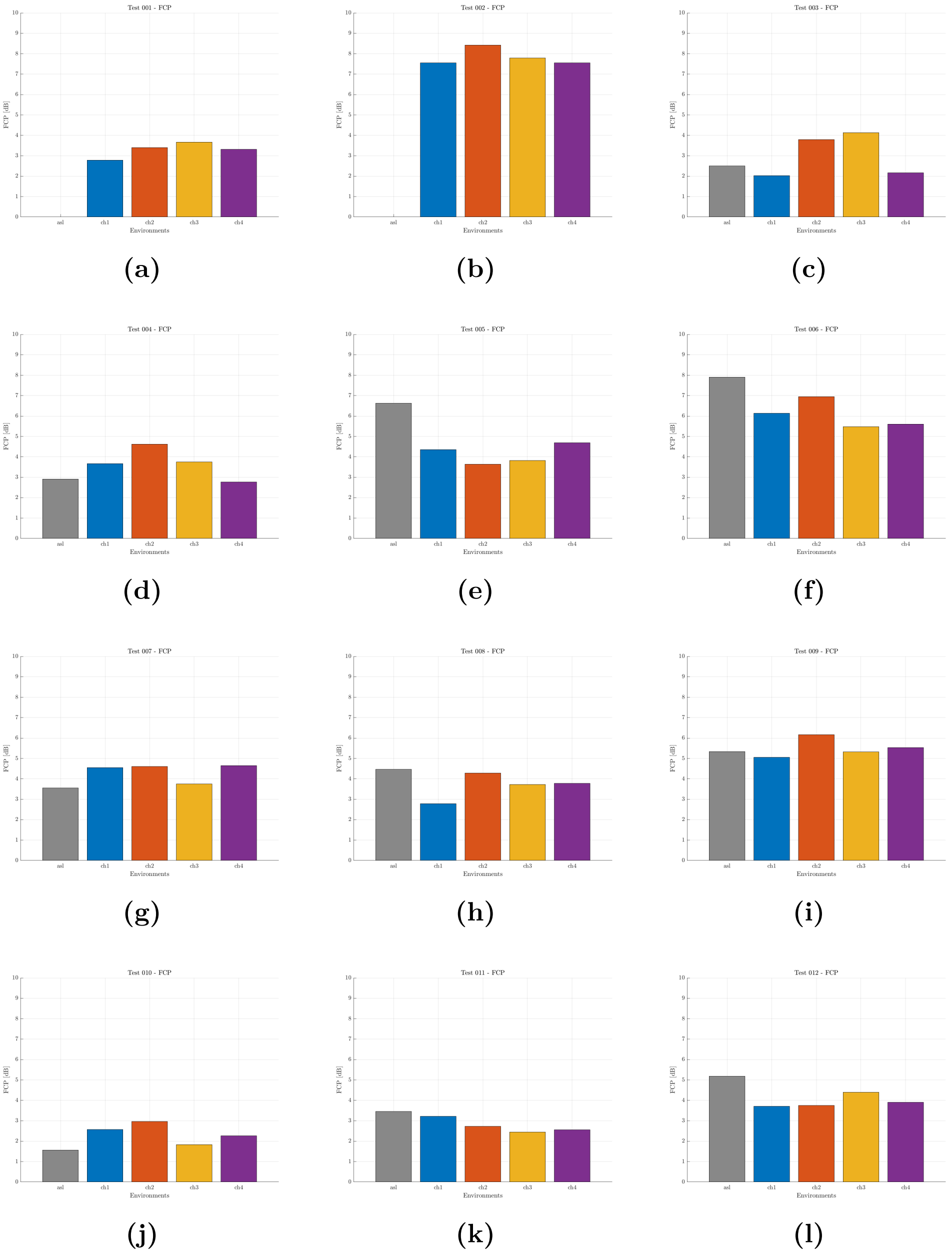


Figure 5.19: Variation of SPR: mean and standard deviation

FCP was computed as shown in Formula 2.23. Values for each volunteer are presented in **Figure 5.20**. Clear differences can be observed depending on the performing environment, although no common trend is immediately apparent. **Figure 5.21** presents the differences in FCP values averaged among the recordings in each of the five environments. It can be observed that *ch2* exhibits a higher mean value compared to the others, while *ch1*, *ch3*, and *ch4* show similar values.

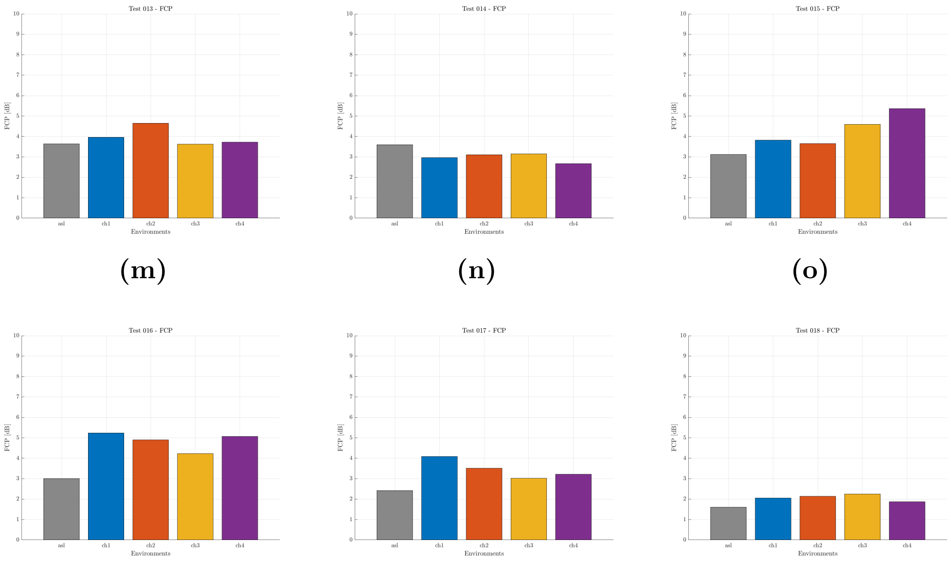


(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g) **(h)** **(i)**

(j) **(k)** **(l)**



(m) **(n)** **(o)**

(p) **(q)** **(r)**

Figure 5.20: Subjective Investigation: FCP

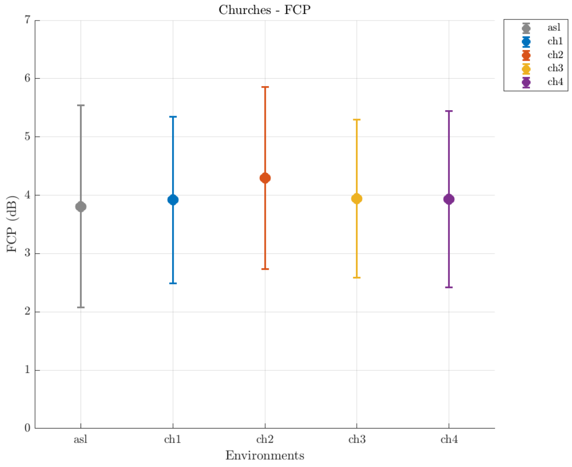
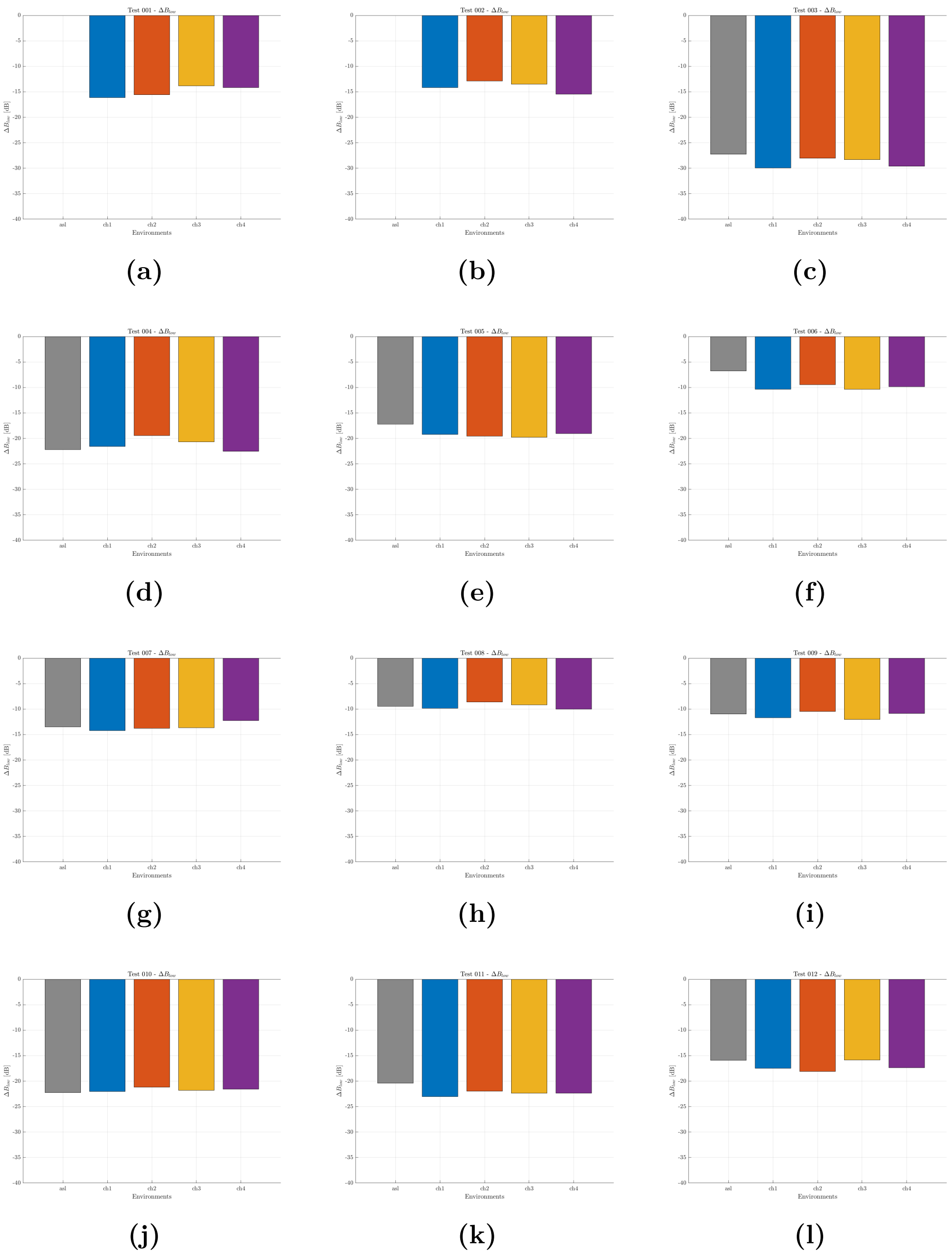


Figure 5.21: Variation of FCP: mean and standard deviation

ΔB_{low} and ΔB_{high} have been computed as a measure of energy balancing of the signal based on its LTAS. Value for each singer are reported in **Figures 5.22** and **5.24**, respectively. Slight differences among the five environments for each of the subjects are noticeable for both indexes.
Figures 5.23 and **5.25** show the mean values among all the volunteers for the five environments: in this graphs, no environment-dependant variations are noticeable.

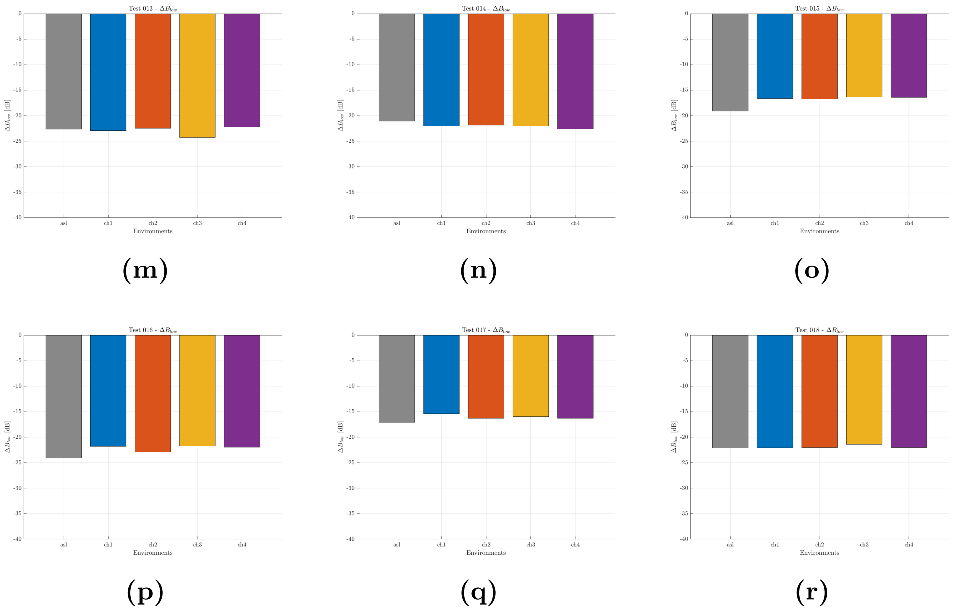


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)



(m) **(n)** **(o)**

(p) **(q)** **(r)**

Figure 5.22: Subjective Investigation: ΔB_{low}

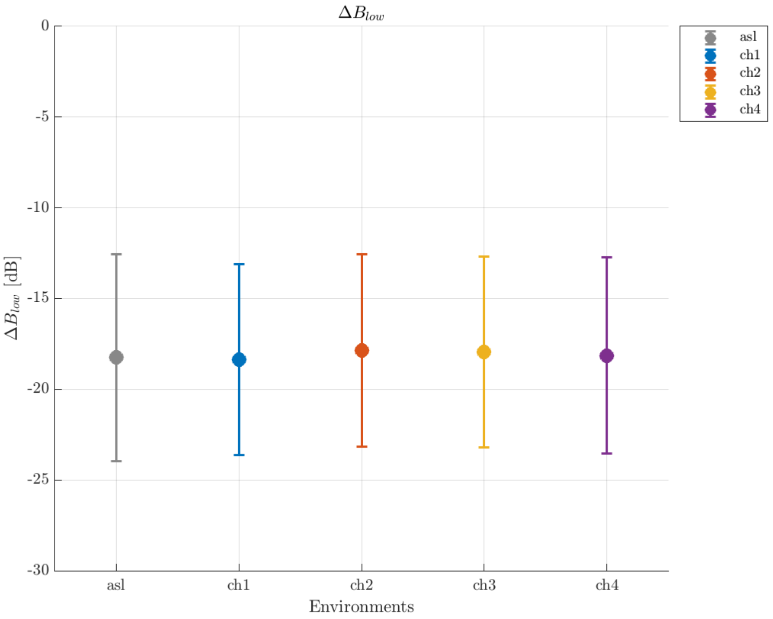
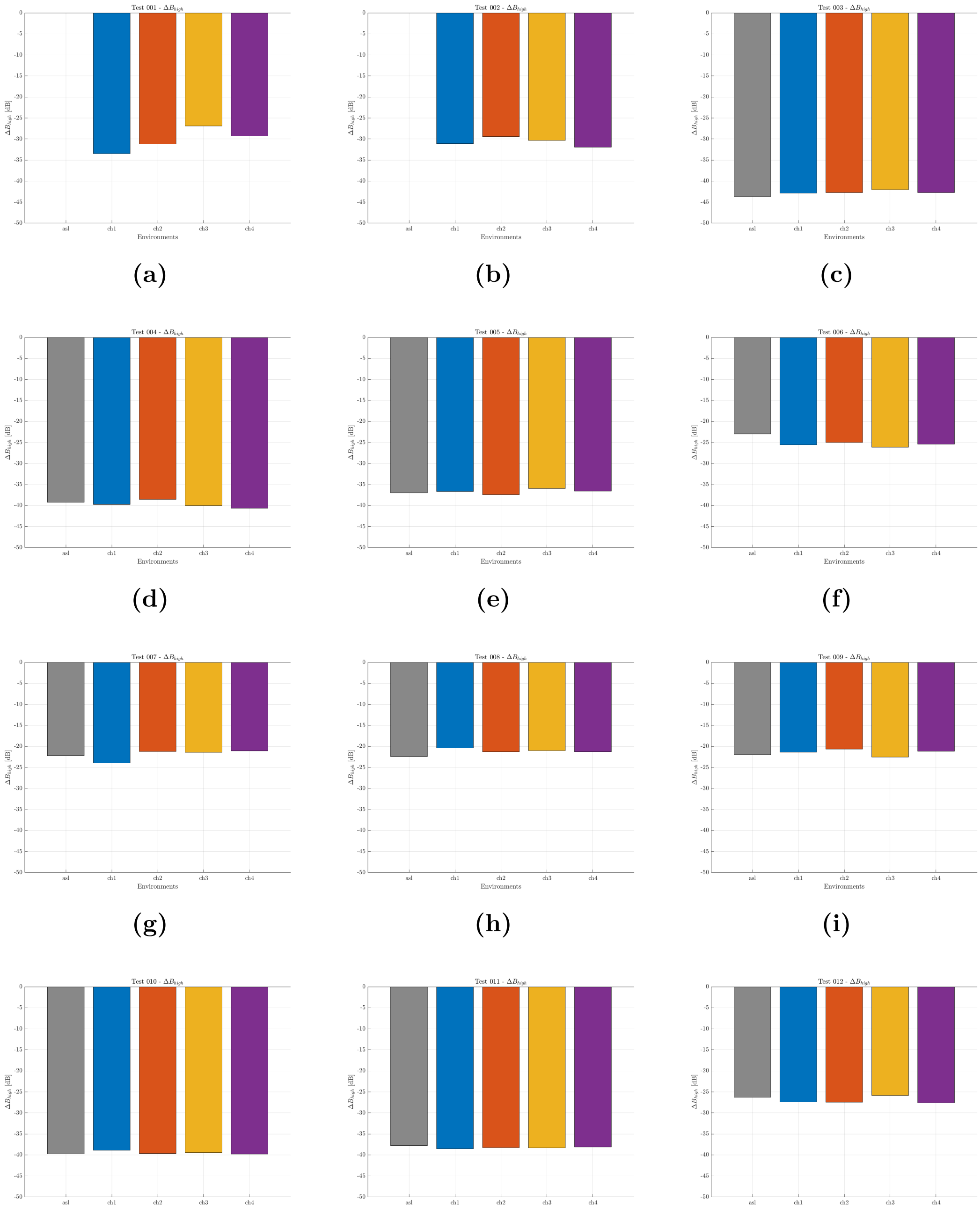


Figure 5.23: Variation of ΔB_{low}: mean and standard deviation

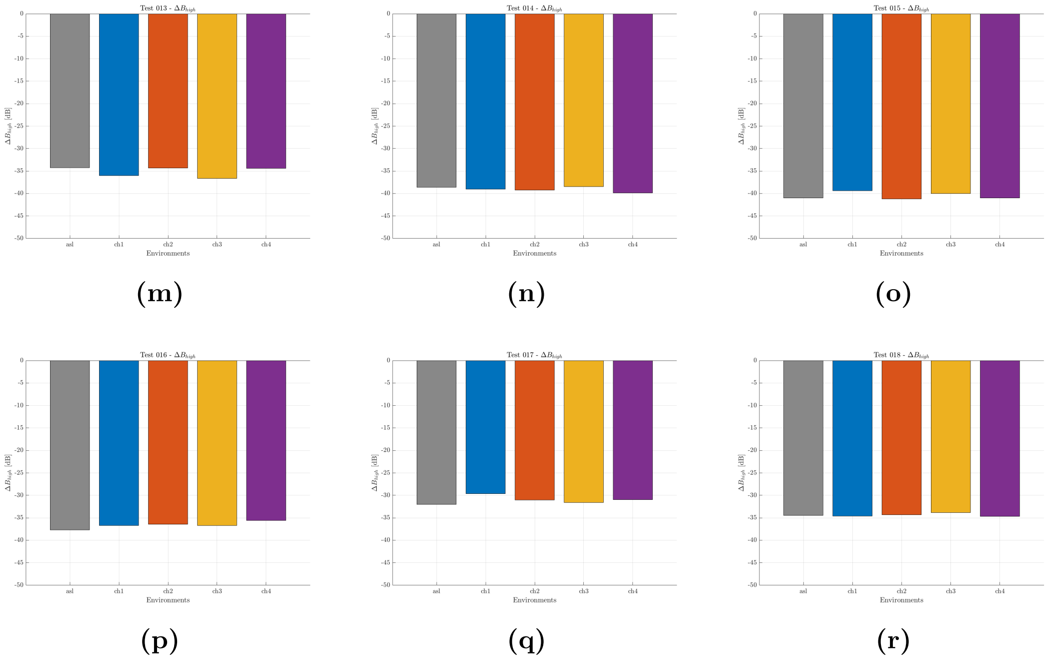


(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g) **(h)** **(i)**

(j) **(k)** **(l)**



(m) **(n)** **(o)**

(p) **(q)** **(r)**

Figure 5.24: Subjective Investigation: ΔB_{high}

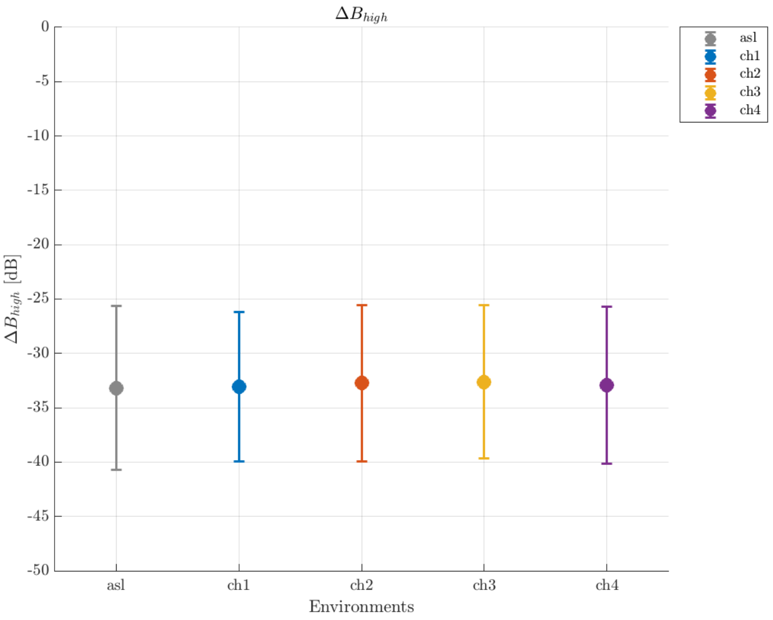
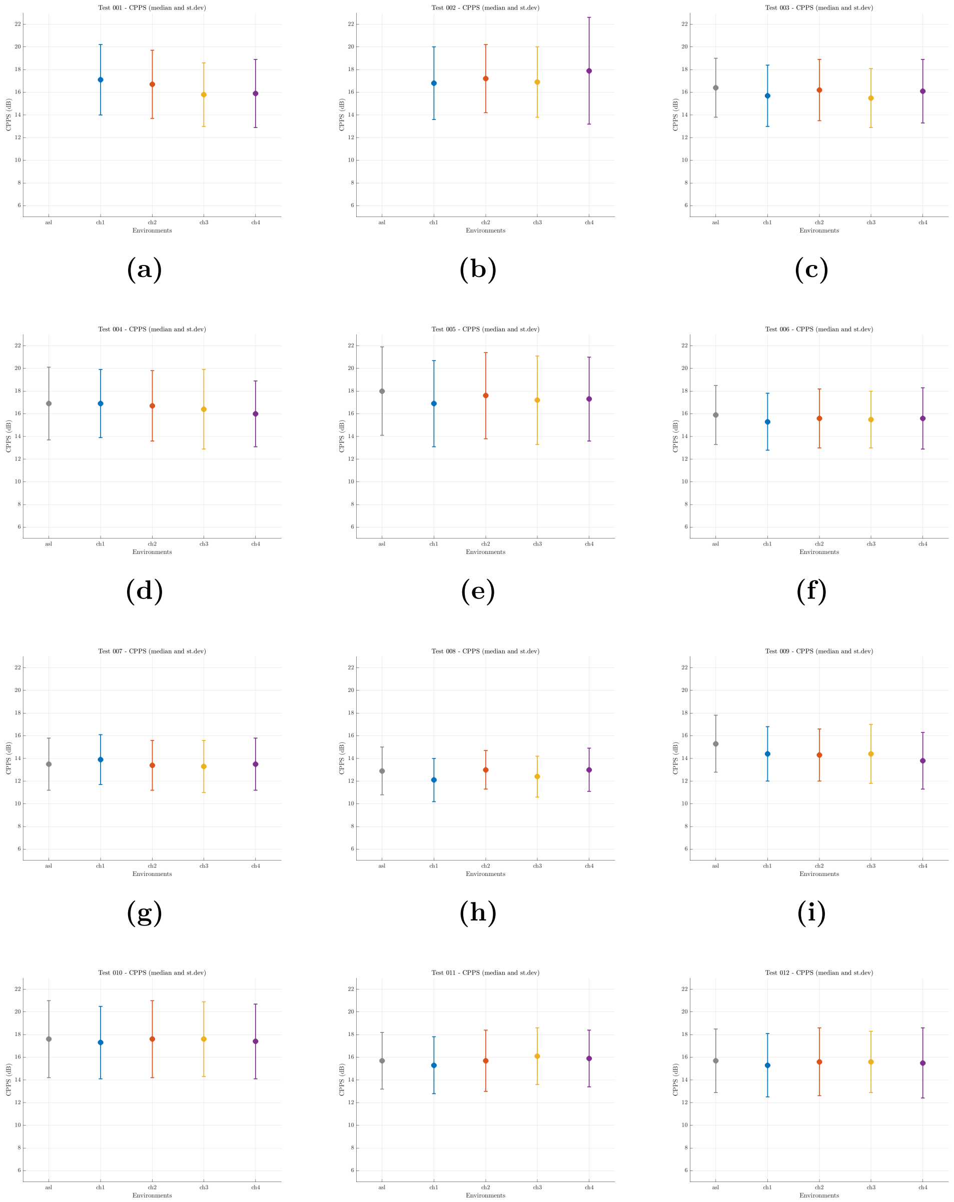


Figure 5.25: Variation of ΔB_{high}: mean and standard deviation

CPPS values were also computed for each recording. **Figure 5.26** presents the median and standard deviation of CPPS for each subject, showing no clear environment-dependent differences. **Figure 5.27** further confirms that the mean values across different participants within the same environment do not vary significantly among the five auralized environments.

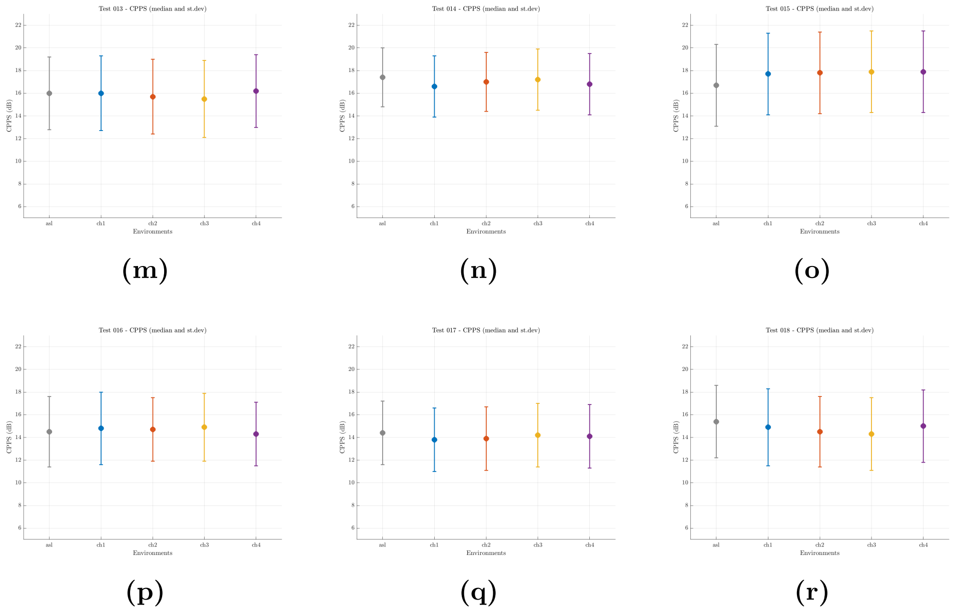


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)



(m) (n) (o)

(p) (q) (r)

Figure 5.26: Subjective Investigation: CPPS

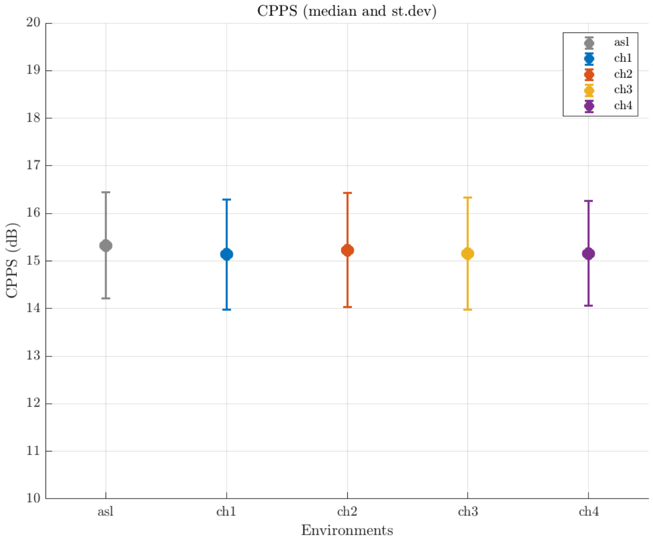


Figure 5.27: Variation of CPPS: median and standard deviation

5.4.2 Subjective Questionnaires

Finally, the results from the subjective questionnaire are presented in **Figure 5.28**. The mean and standard deviation of participants' responses are shown, maintaining the original order of the questions as they appeared in the questionnaire. Noticeable differences can be observed for the following indices: Three-Dimensionality, Presence, Spatiality, and Reverberation.

The values for Plausibility and Ease of Singing indicate a generally positive evaluation across all environments, which can be considered a good validation of the auralization output. In contrast, the indices for Coloration, Clarity, Pitch, Vibrato, and Velocity are all centered around the neutral position. This could suggest either that these aspects were not strongly perceived by the participants or that the definitions provided in the questionnaire lacked clarity.

Lastly, the Preference of the participants is reported, showing that *ch2* was the most favourite among all, while the other three environments were rated as equally prefered on average. This ranking partially confirms the feedback given by the professional singers, although *ch4* was evaluated as less lickable from the amateurs. Investigation on the reasons of this gap may be the subject of future researches on the differences between the auralized and actual environments, and how they may influence the lickability perceived by the users.

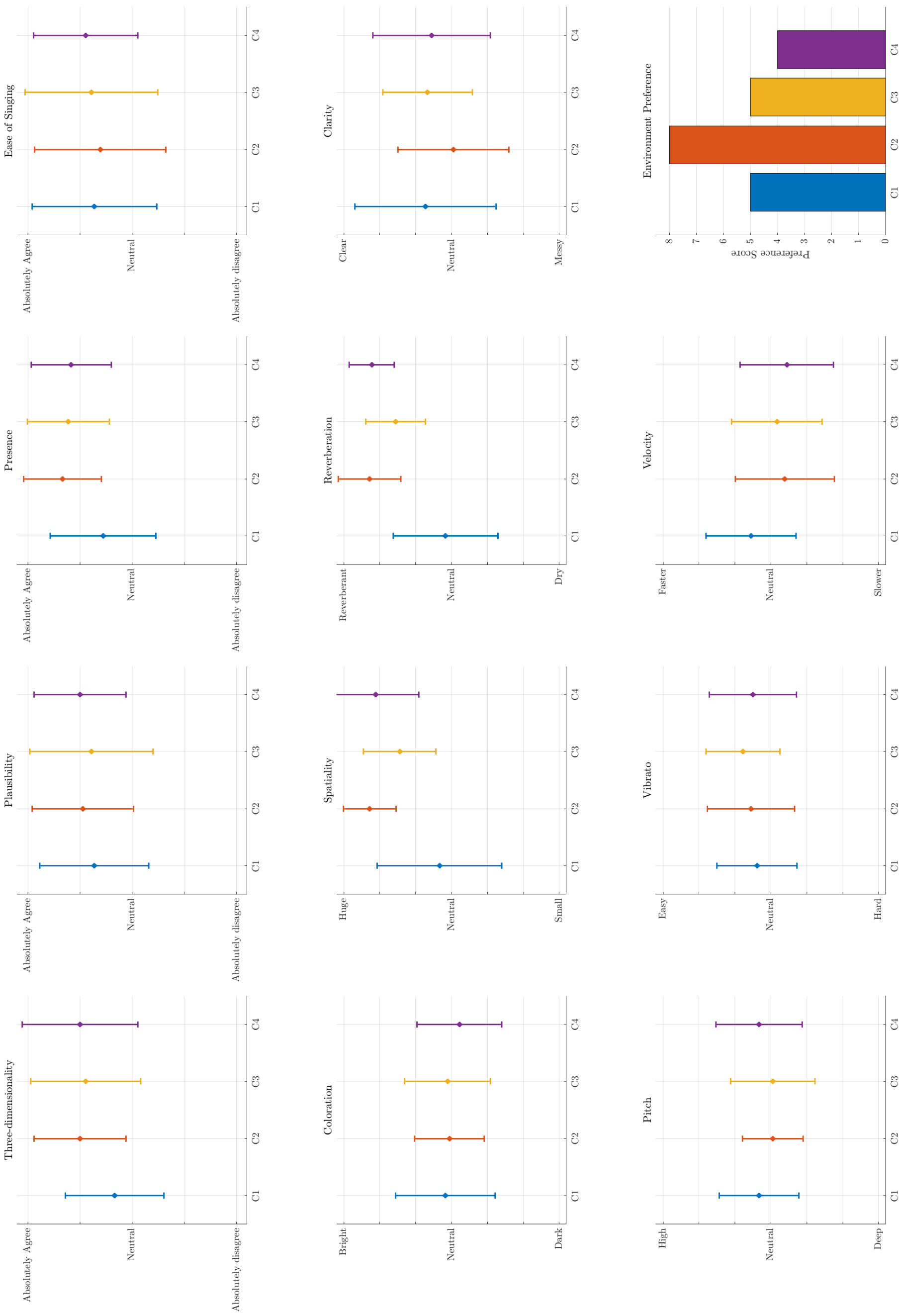


Figure 5.28: Subjective questionnaire responses: mean and standard deviation

5.4.3 Free Comments on auralization

Comments on the experience in each auralized environment and on the overall testing protocol were collected. Only comments mentioned by more than one participant are reported here.
It is important to note that these remarks come from a group of amateur singers who did not interact with each other and received no bias from the operator of the auralization system. Their opinions are based on a limited or lacking knowledge of both acoustical and music and may have been heavily influenced by their psychological state and general comfort during the entire protocol.

- Church 1
  - The ambient seems big, but the reverberation in not as much as expected.
  - It is uncomfortable to sing in here, because little support is provided
  - It is like a big, empty room (as a gym)
  - If not uncomfortable, it is surely unpleasent to sing in here
- Church 2
  - It seems to be similar to a big empty church or a big empty hall of a castle
  - The sound is enveloping
  - The feedback sometimes is too messy
  - The feedback comes late, forcing the singer to slow down
  - It provides a vertical impact, more than the other environments
- Church 3
  - The late reverberance seems to be pretty brilliant
  - The sound reflects early from the virtual "ceiling", while it disperse on the sides
  - The space seems big but oppressive, as in underground spaces (cellar, cave, tunnel)
  - It would be difficult to imagine a room which naturally sounds like this
  - The feedback is not messy, it is comfortable for singing
- Church 4
  - The entire feedback seems to be brighter than expected
  - There's an emphasys or dissonance that is difficult to identify

  - It is hard to imagine an environment that sounds like this, it seems almost as digitally synthetized

Generally speaking, approximately half of the participants complained that the feedback volume in every environment was too low and that, most of the time, the reverberation was masked by the singer's own emission. However, this feedback concerned only the perceived volume and was reported to have little effect on the realism of the auralization. Some participants also expressed difficulty in finding appropriate terminology to describe their auditory perception in a meaningful way. Additionally, several participants noted that most of the feedback appeared to come from the front direction, with little to no sound perceived from the rear part of the system. This sensation was likely a consequence of the IR measurement procedure in the churches and may be an important aspect to consider in future improvements of the proposed system.
Lastly, two participants found the testing session too long, albeit for different reasons: one considered singing for 40 minutes too demanding, while the other pointed out that their voice gradually warmed up throughout the session, making the latter part of the test naturally easier. Future in-depth analyses could further investigate the importance of vocal warm-up prior to such subjective evaluation experiences.

5.5 Soloist in ASL

The same experimental protocol in the ASL was conducted with the same soloist who performed in the actual churches. As with the amateur singers, the soloist was instructed to perform freely, without restrictions on repertoire, tempo, or pitch. During the third block of the testing procedure, the same song (“Happy Birthday to You”) was recorded, with the same reference note provided prior to each take. The vocal parameters extracted from these recordings are presented in the following sections.

The SPL during the tests was calculated using the calibrated head-worn microphone, following the same procedure detailed for the amateur singers. As shown in **Figure 5.29**, the mean level across all environments ranged from 85 to 90 dB, which is notably higher than the mean level observed among the amateur singers. No significant variation in this parameter is evident across the five environments.

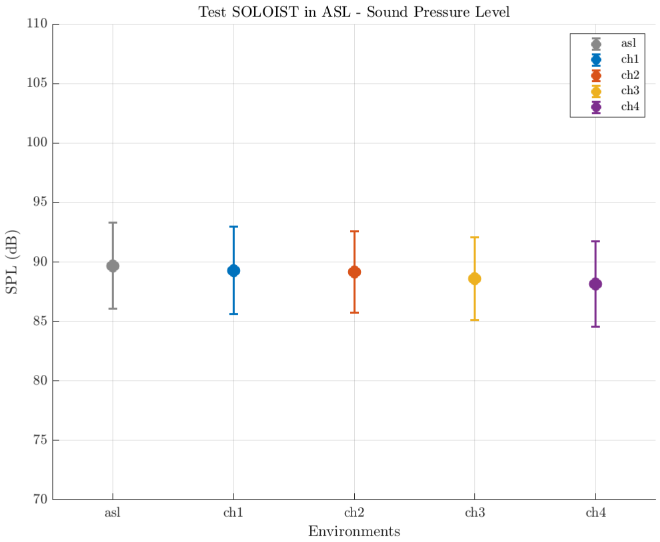


Figure 5.29: Variation of SPL: mean and standard deviation

TD and Pause Length for the soloist across the different environments are presented in **Figure 5.30** and **Figure 5.31**, respectively. While TD exhibits a slight variation, the length of the pauses differed significantly across the five recordings.

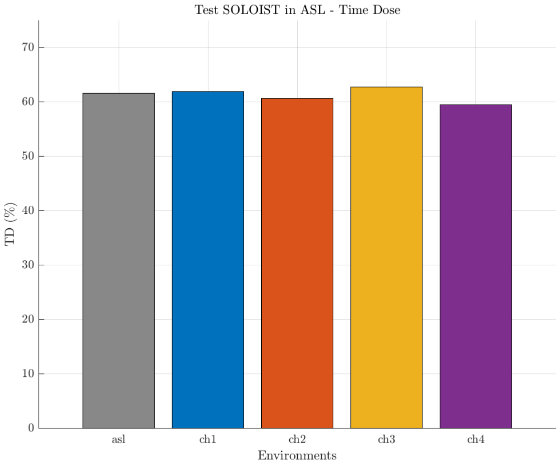


Figure 5.30: Variation of Time Dose: mean and standard deviation

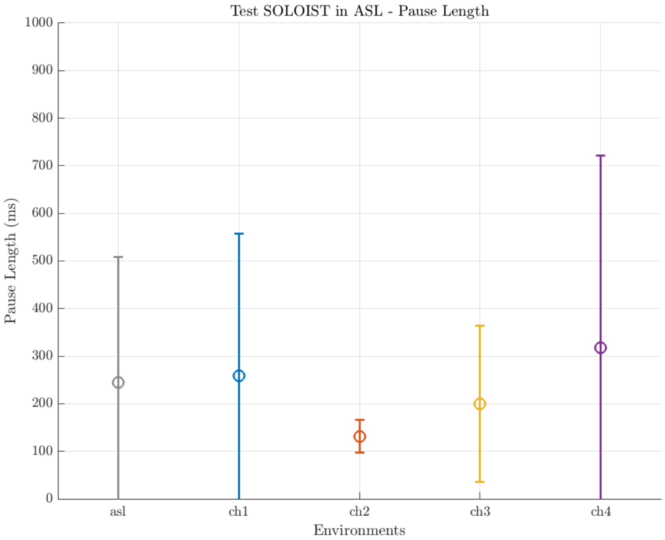


Figure 5.31: Variation of Pause Length: mean and standard deviation

The distribution of the mean F_0 across the five environments is presented in **Figure 5.32**. The distribution curve remains consistent regardless of the environment being simulated in the ASL. A similar trend is observed in **Figure 5.33**.

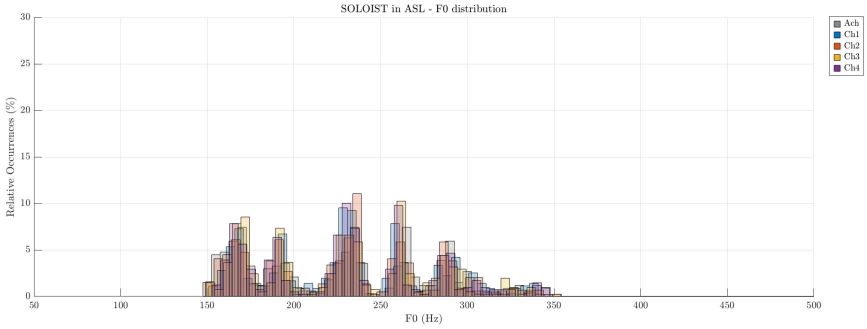


Figure 5.32: Distribution of F0 - Soloist in ASL

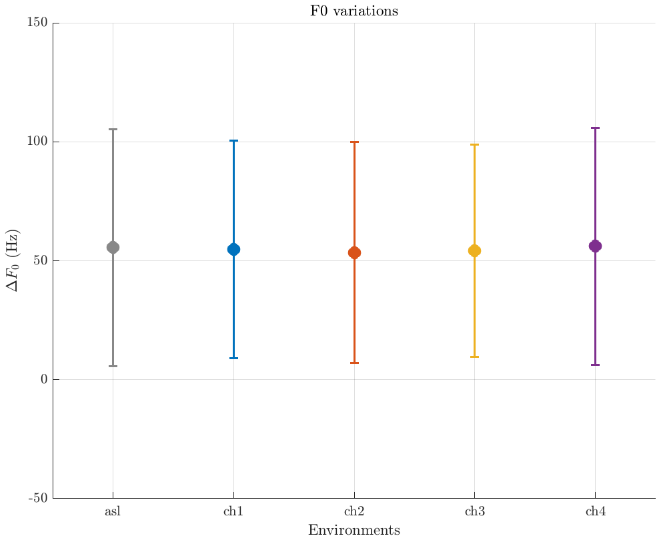


Figure 5.33: Variation of F0 to the given note: mean and standard deviation

The LTAS of the five recordings is depicted in **Figure 5.34**. Similar to the results observed for the amateur singers, the curves align closely in the low-frequency range but exhibit significant differences above 1500 Hz. No evident single formant cluster is observable in the range from 2000 Hz to 4000 Hz, were the Singers' Formant should be found.

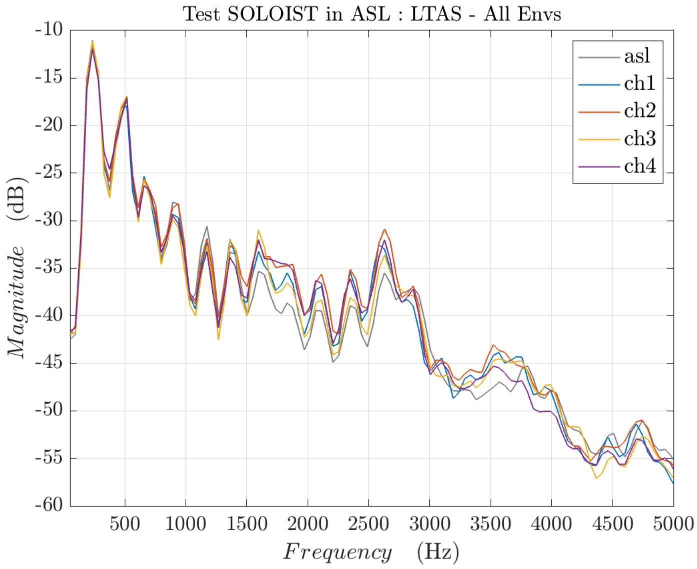


Figure 5.34: LTAS of Soloist during test

A set of features was extracted from the LTAS, employing the same methodology utilized for the amateur singers. These features are summarized in Table 5.6. The variations across the different environments are generally negligible, with the exception of FCP, which demonstrates a more pronounced difference.

Parameter	**Asl**	**Ch 1**	**Ch 2**	**Ch 3**	**Ch 4**
SPR [dB]	-21.6	-20.7	-20.0	-21.1	-21.0
FCP [dB]	5.4	7.1	7.6	6.3	7.1
ΔB_{low} [dB]	-13.3	-12.8	-11.9	-13.2	-12.2
ΔB_{high} [dB]	-29.3	-31.2	-30.7	-31.1	-30.9

Table 5.6: Soloist voice parameters during test: LTAS features

Lastly, the CPPS values derived from the recordings are presented in **Figure 5.35**. This figure illustrates an almost identical distribution of values across the environments, with the sole exception of the recording in environment *ch3*, which exhibits a marginally higher median value. A similar observation can be made in **Figure 5.36**.

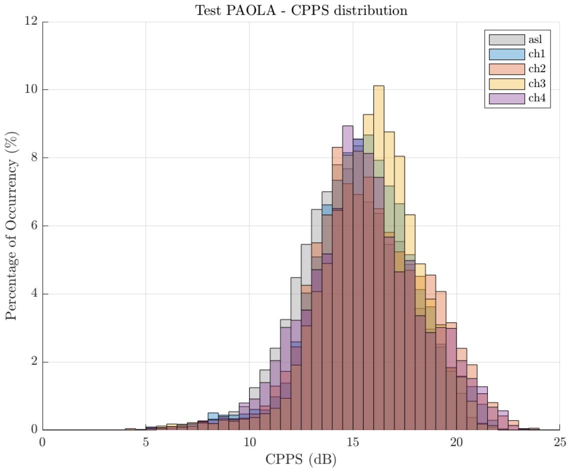


Figure 5.35: Distribution of CPPS: median and standard deviation

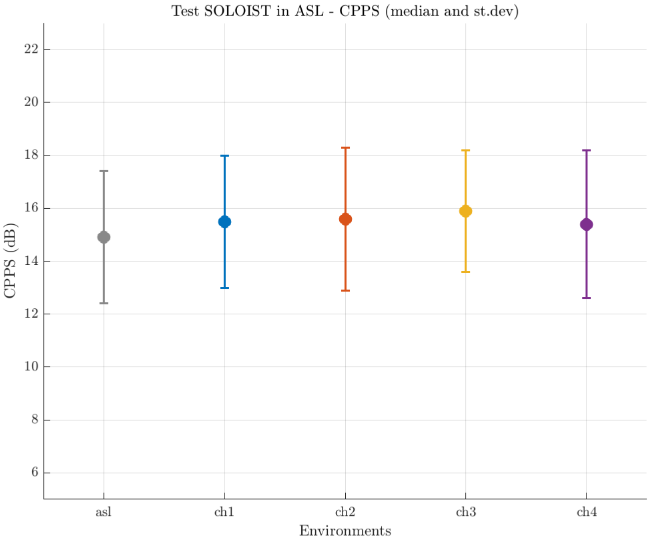


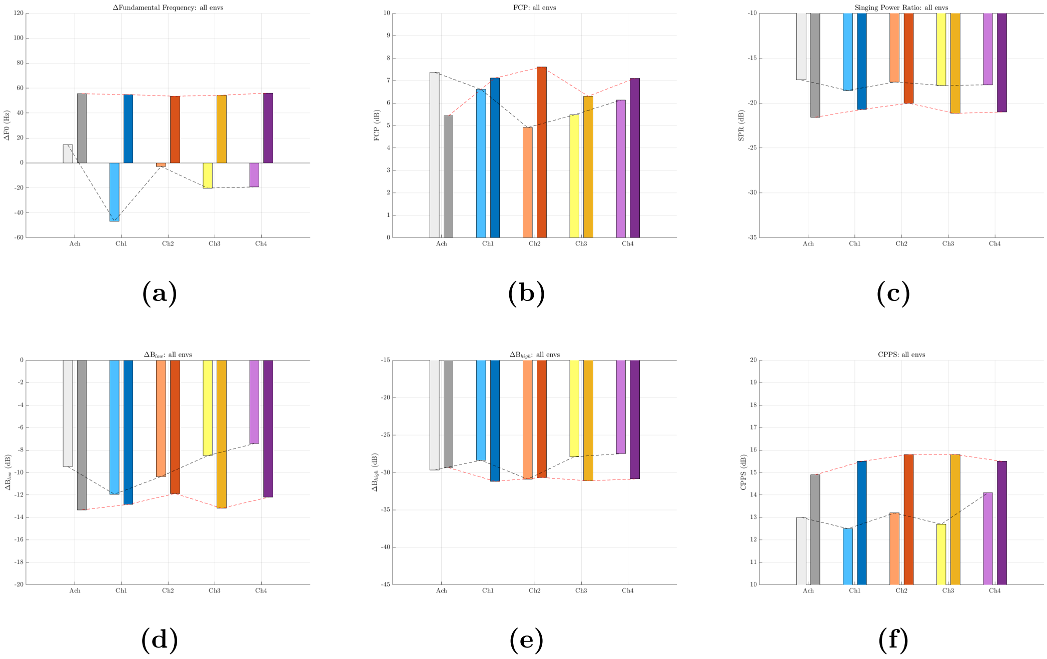
Figure 5.36: Variation of CPPS: median and standard deviation

5.6 Comparison of Soloist Voice Parameters between Real Churches and Auralized Environments

Figure 5.37 presents a comparison of the same vocal parameters extracted from the soloist performing both in actual churches *(on the left, brighter colors)* and in the auralized environments in the ASL *(on the right, darker colors)*. For the comparison with the anechoic case, the recordings performed in the ASL with the auralization engine turned off have been considered and labeled as *Ach*.

A similar trend can be observed in the variation of SPR across the environments: in both datasets, the value of this parameter is highest for Church 2, while it is slightly lower in the other three churches. The value computed in the anechoic chamber is significantly higher than in the real Church 2, whereas the value from the recording in the bare ASL is significantly lower than the value obtained from the auralization using the modified IR from Church 2.

All the other parameters show no consistent patterns of variation between the real churches and the auralized environments.



(a) (b) (c)

(d) (e) (f)

Figure 5.37: Voice Parameters of Soloist: churches *(on the left, brighter colors)* versus ASL *(on the right, darker colors)*

5.7 Acoustical Parameters versus Voice Parameters

To verify whether there was a correlation between the acoustical parameters and the vocal parameters analyzed in this thesis, a series of graphs was prepared and is reported in the following pages. The plots present the voice parameters as dependent variables of the acoustical parameters, which are considered independent variables.

This analysis was conducted for both the vocal parameters of the soloist recorded in the churches and those extracted from the recordings in the ASL. Additionally, the same procedure was applied to the subjective impressions collected through the questionnaire. The results are presented and discussed in the following sections.

For indexing purposes, the acoustical parameters recorded in the real churches are labeled with a subscript *c*, resulting in parameters such as EDT_c, $\mathrm{T}_{20,c}$, $\mathrm{C}_{80,c}$, $\mathrm{D}_{50,c}$, $\mathrm{T}_{s,c}$, BR_c, and TR_c. Conversely, parameters measured in the auralized environments are denoted with the subscript *asl*, yielding parameters such as EDT_{asl}, $\mathrm{T}_{20,asl}$, and $\mathrm{DT}_{40,ME,asl}$.

The color coding used in the following sections in the same as used in previous graphs, but for the sake of clarity, the legend is reported below.

🟦 Church 1
🟧 Church 2
🟨 Church 3
🟪 Church 4

5.7.1 Acoustics versus Soloist Voice Parameters in Churches

Figures 5.38 presents the correlation between acoustical parameters and the soloist voice parameters.

The variation of the mean F_0 (named ΔF_0) shows a linear relation with all the indicators of Reverberation Time - in particular, EDT$_c$, T$_{20,c}$, EDT$_{asl}$, T$_{20,asl}$, and DT$_{40,ME,asl}$, but also with T$_s, c$. Both C$_{80,c}$, D$_{50,c}$ show an opposite relation with this parameter, as this indexes decrease the more energy is concentrated in the later part of the IR.

The other plots do not show any observable relation between voice and acoustical parameters, although the graphs for T$_{20,asl}$ and EDT$_c$ show superimposable behaviours for every vocal parameter.

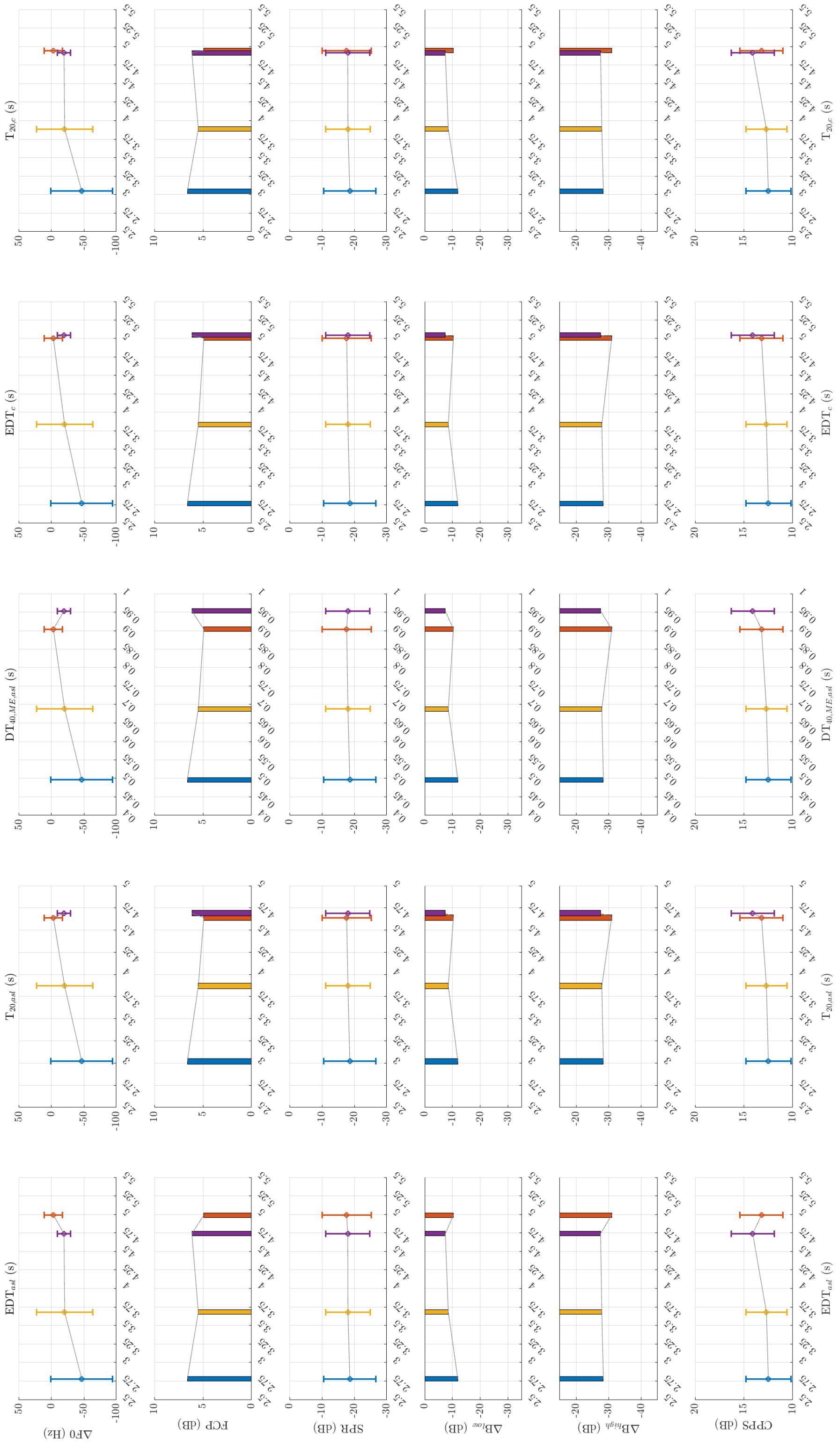
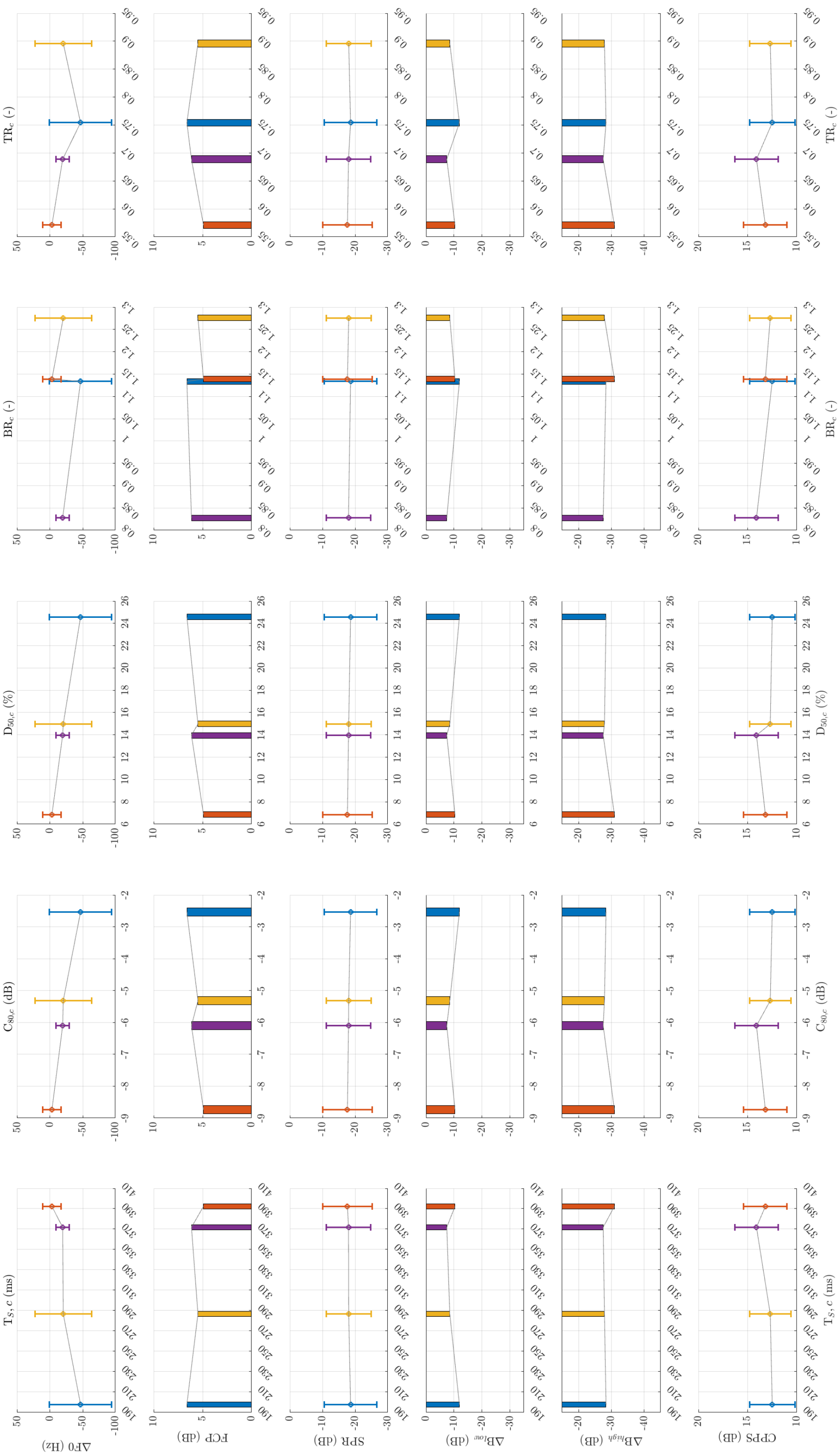


Figure 5.38: Acoustical Parameters compared to Soloist voice parameters (1/2)



Acoustical Parameters compared to Soloist voice parameters (2/2)

5.7.2 Acoustics versus Voice Parameters from Tests in ASL

Figure 5.39 shows the relations between acoustical parameters and vocal parameters extracted from the recording performed during the subjective investigation in the ASL.

TD shows a linear descending correlation with $\mathrm{T}_{20,asl}$, $\mathrm{DT}_{40,ME,asl}$ EDT_c. The correlation with the other acoustical parameters related to the evaluation of the reverberation time yield similar behaviours.

CPPS shows similar linear progression compared to the indexes of reverberation time, in particular EDT_{asl} and $\mathrm{DT}_{40,ME,asl}$, but the differences are too little to allow for a certain statement on this correlation.

The other voice parameters do not show any relevant variations across the different environments, so no significant relations with the acoustical parameters may be observed.

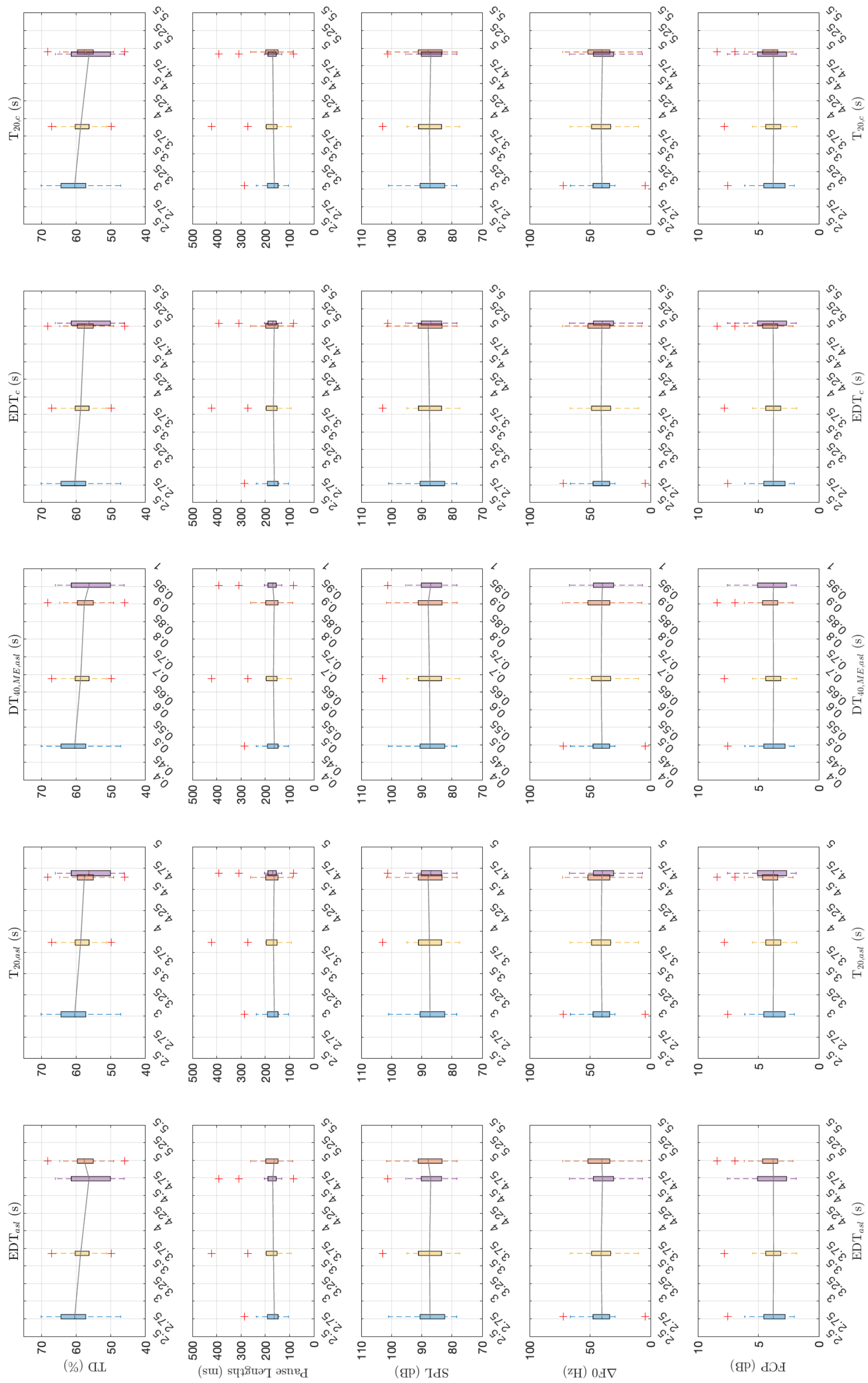
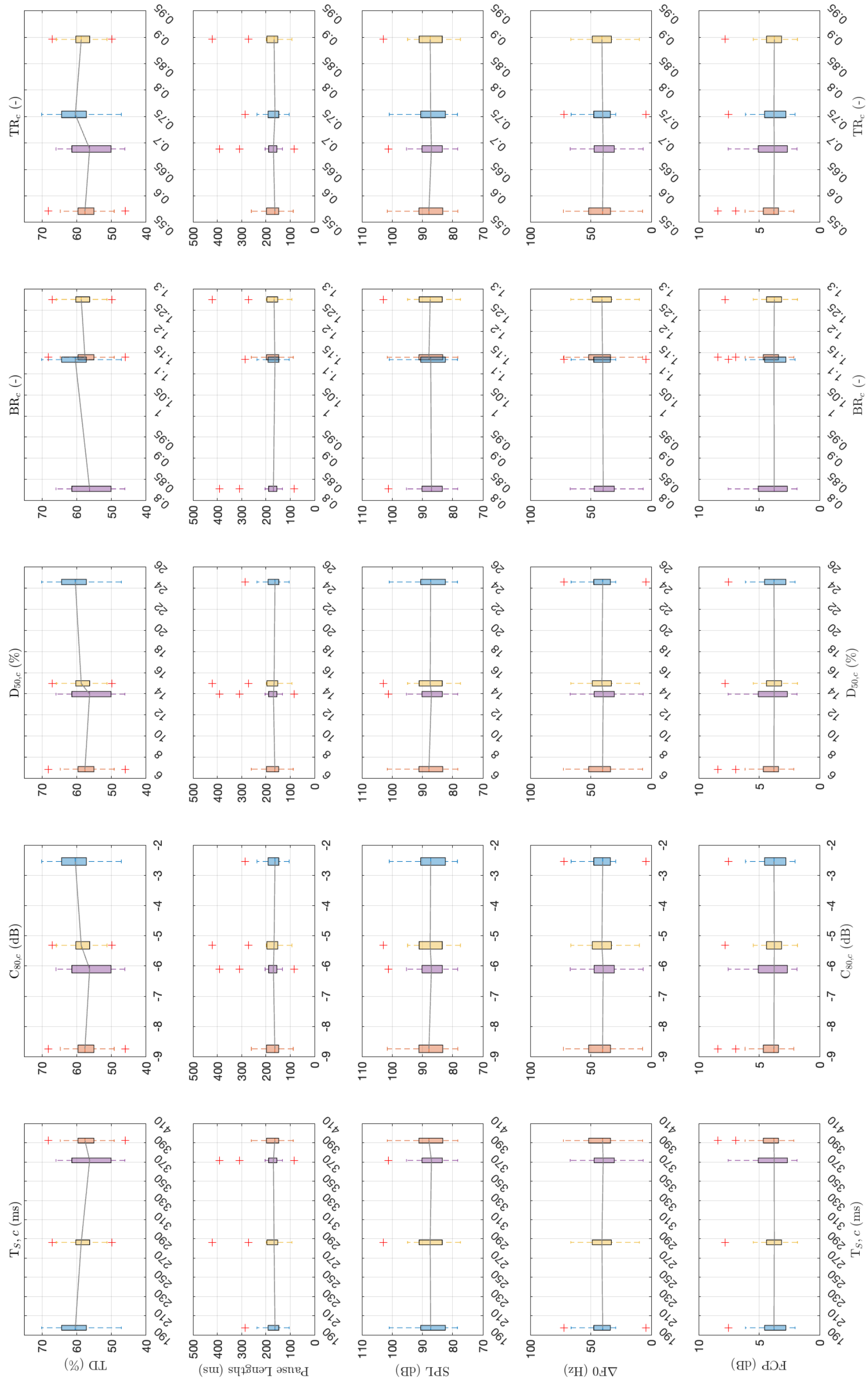
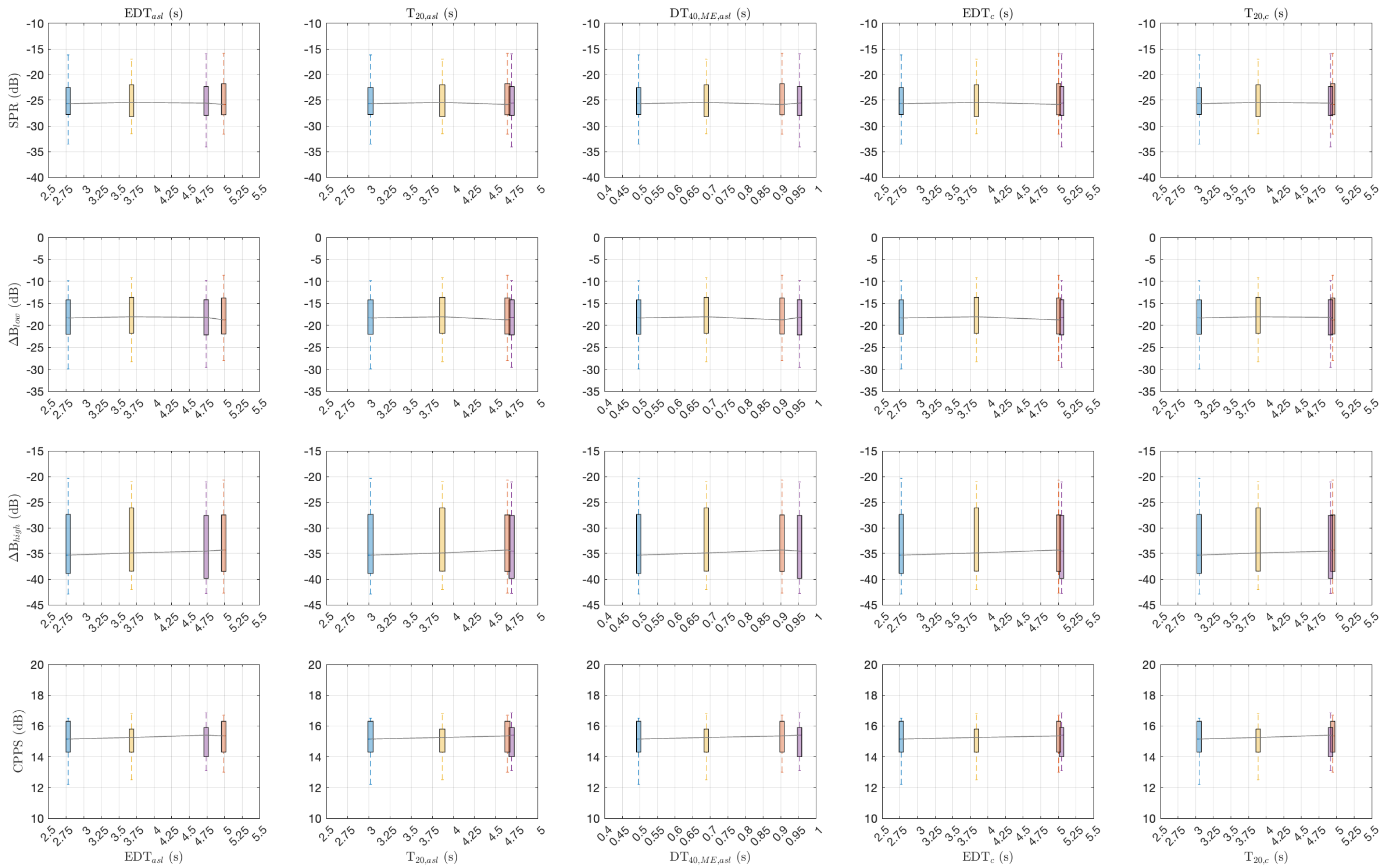


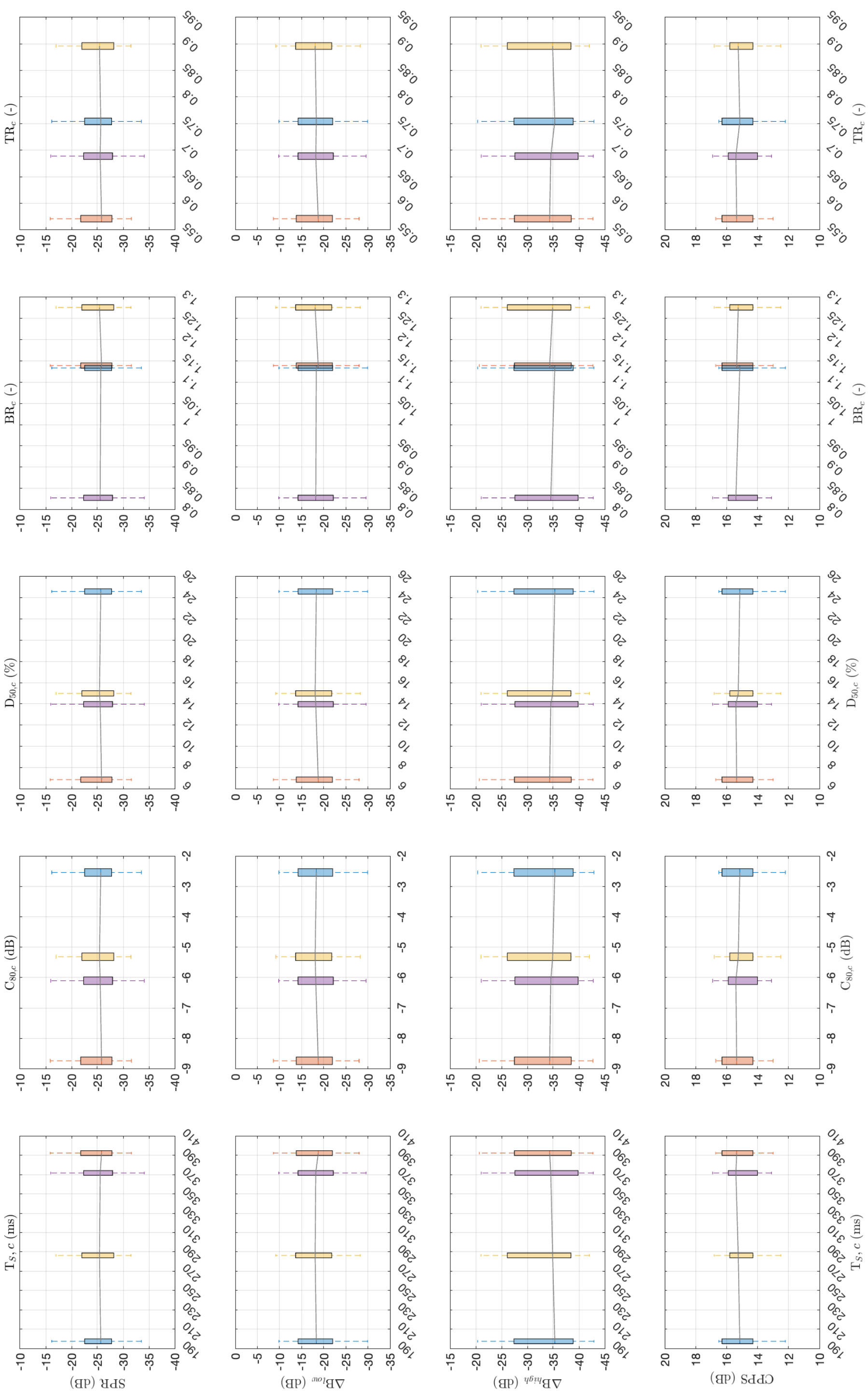
Figure 5.39: Recording in ASL versus Acoustical Parameters (1/4)



Recording in ASL versus Acoustical Parameters (2/4)



Recording in ASL versus Acoustical Parameters (3/4)



Recording in ASL versus Acoustical Parameters (4/4)

5.7.3 Acoustics versus Questionnaire in ASL

Figure 5.40 presents the plots of the questionnaire responses correlated with the acoustical parameters.

Many observation may be collected from this figures.

The sense of Three-Dimensionality shows an ascending trend, which correlates this perception with all the Reverberation Time indexes—EDT$_{asl}$, T$_{20,asl}$, DT$_{40,ME,asl}$, EDT$_c$, and T$_{20,c}$—while displaying a negative linear trend when compared to C$_{80,c}$ and D$_{50,c}$.

The sense of Presence shows a similar ascending trend, particularly for EDT$_{asl}$ and T$_{s,c}$, while the other parameters follow a similar ranking of values, except for those recorded for Church 4 (represented by the purple bar).

The senses of Reverberation and Spatiality exhibit a strong correlation with all of the Reverberation Time parameters (EDT$_{asl}$, T$_{20,asl}$, DT$_{40,ME,asl}$, EDT$_c$, and T$_{20,c}$) and with T$_{s,c}$, while both senses show a negative correlation with C$_{80,c}$ and D$_{50,c}$.

The sense of Clarity shows slight correlations with EDT$_{asl}$, C$_{80,c}$, D$_{50,c}$, and TR$_c$, but statements regarding these relations may not be meaningful due to the small variance of this perceptual parameter reported across the different auralized environments, and the large standard deviations in the responses from different participants regarding the same environment.

The issues on the meaningfulness of the senses of Pitch, Vibrato and Velocity have already been discussed in the previous sections. Nevertheless, the sense of Velocity (intended as the tendency of the environment to affect the singer's performance) show a relevant negative correlation with all the indexes correlated to the Reverberation Time—particularly T$_{20,asl}$, DT$_{40,ME,asl}$, EDT$_c$, and T$_{20,c}$— and with T$_{s,c}$, while it shows a slight positive correlation with C$_{80,c}$ and D$_{50,c}$.

Generally speaking, in the majority of cases a similar behaviour may be observed on all the perceptual parameters while considering EDT$_{asl}$, T$_{20,asl}$, DT$_{40,ME,asl}$, EDT$_c$, T$_{20,c}$ and T$_{s,c}$. These trends are also widely in counterposition with the trends showed by C$_{80,c}$ and D$_{50,c}$.

These findings can be interpreted as a general positive feedback on the auralization engine, as the parameters collected in real environments and those measured in auralized situations exhibit very similar trends. This suggests that the results from both contexts can be meaningfully compared. Prior to these findings, it was not possible to ensure this comparability, as the system had not been validated using a standardized, unequivocal method.

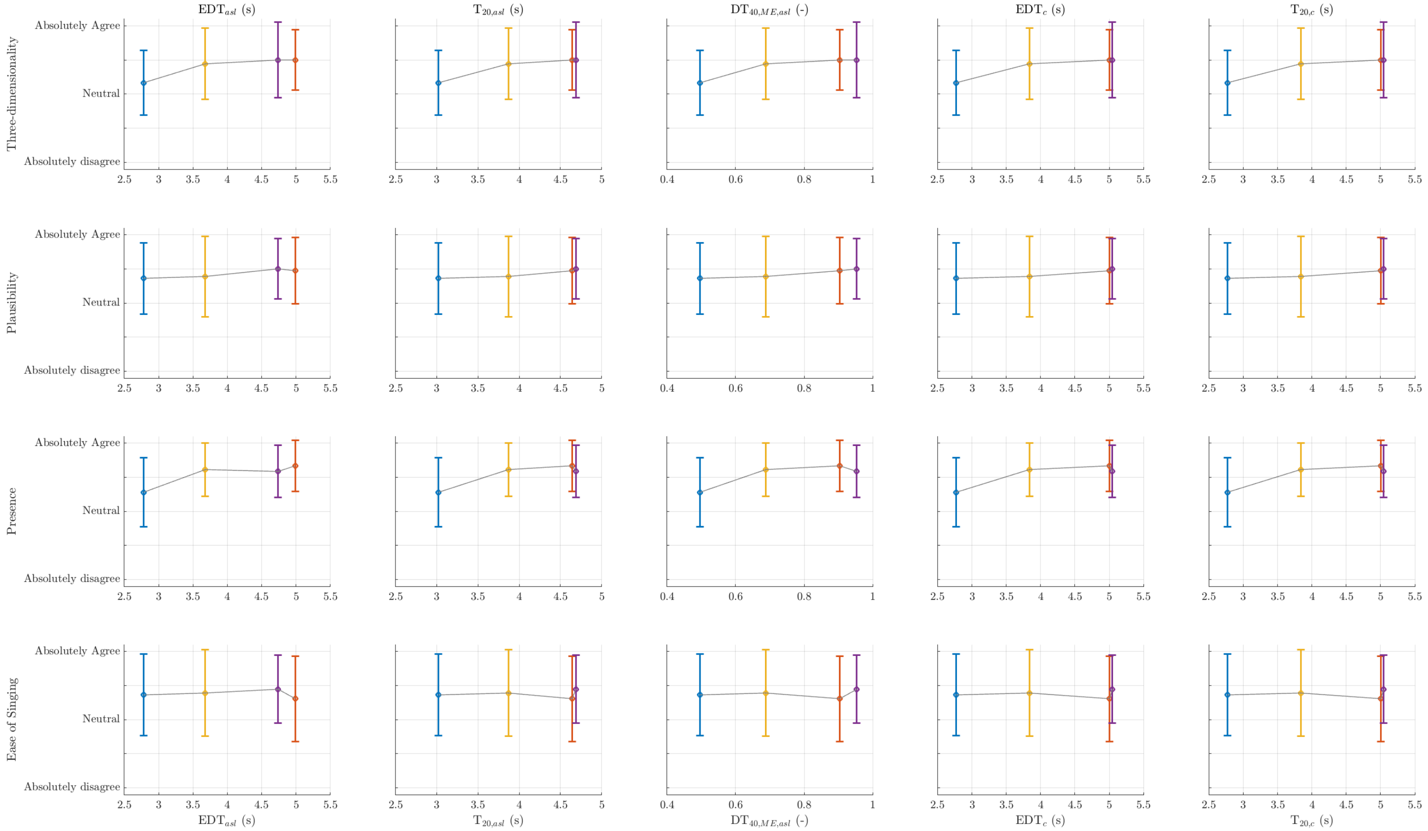
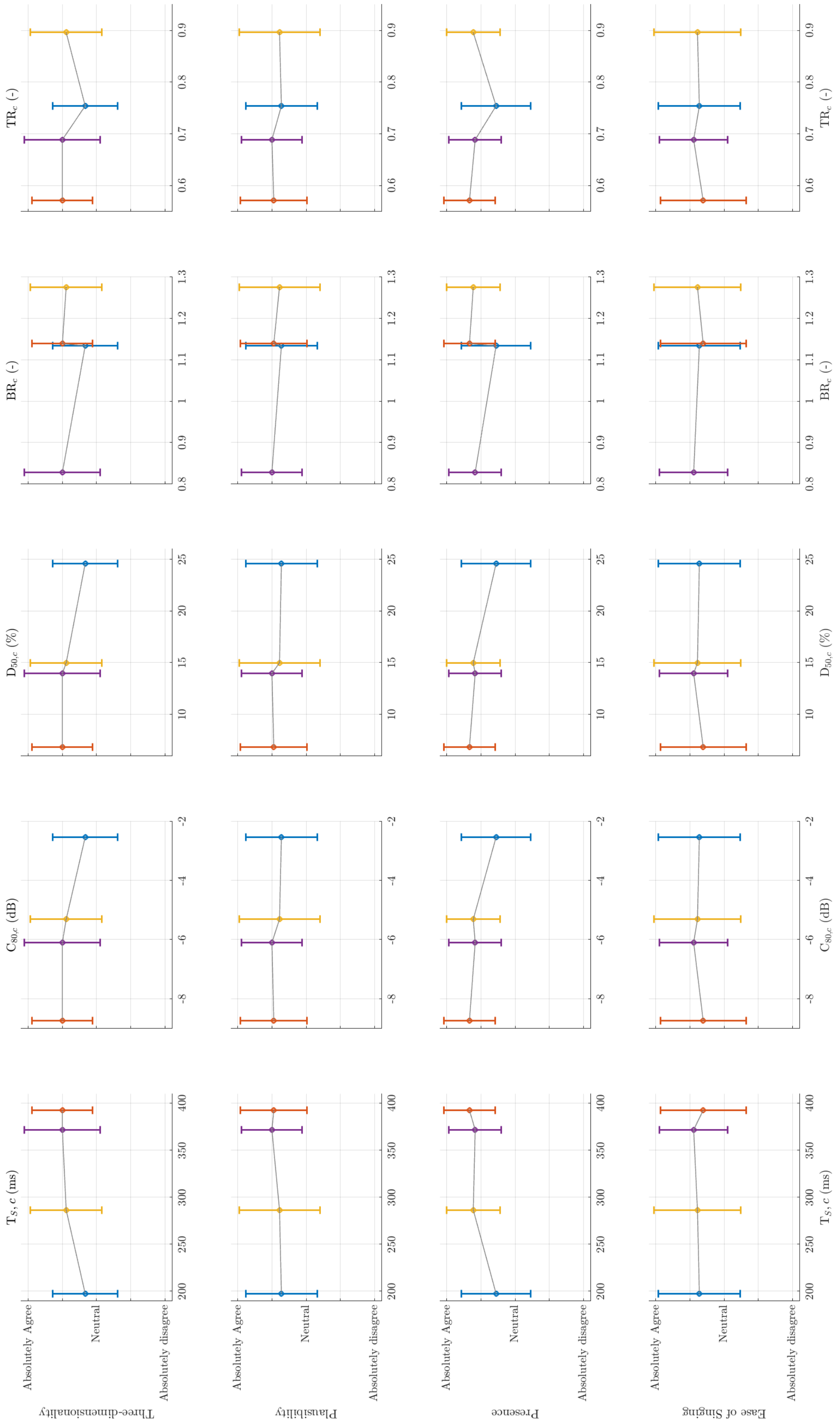
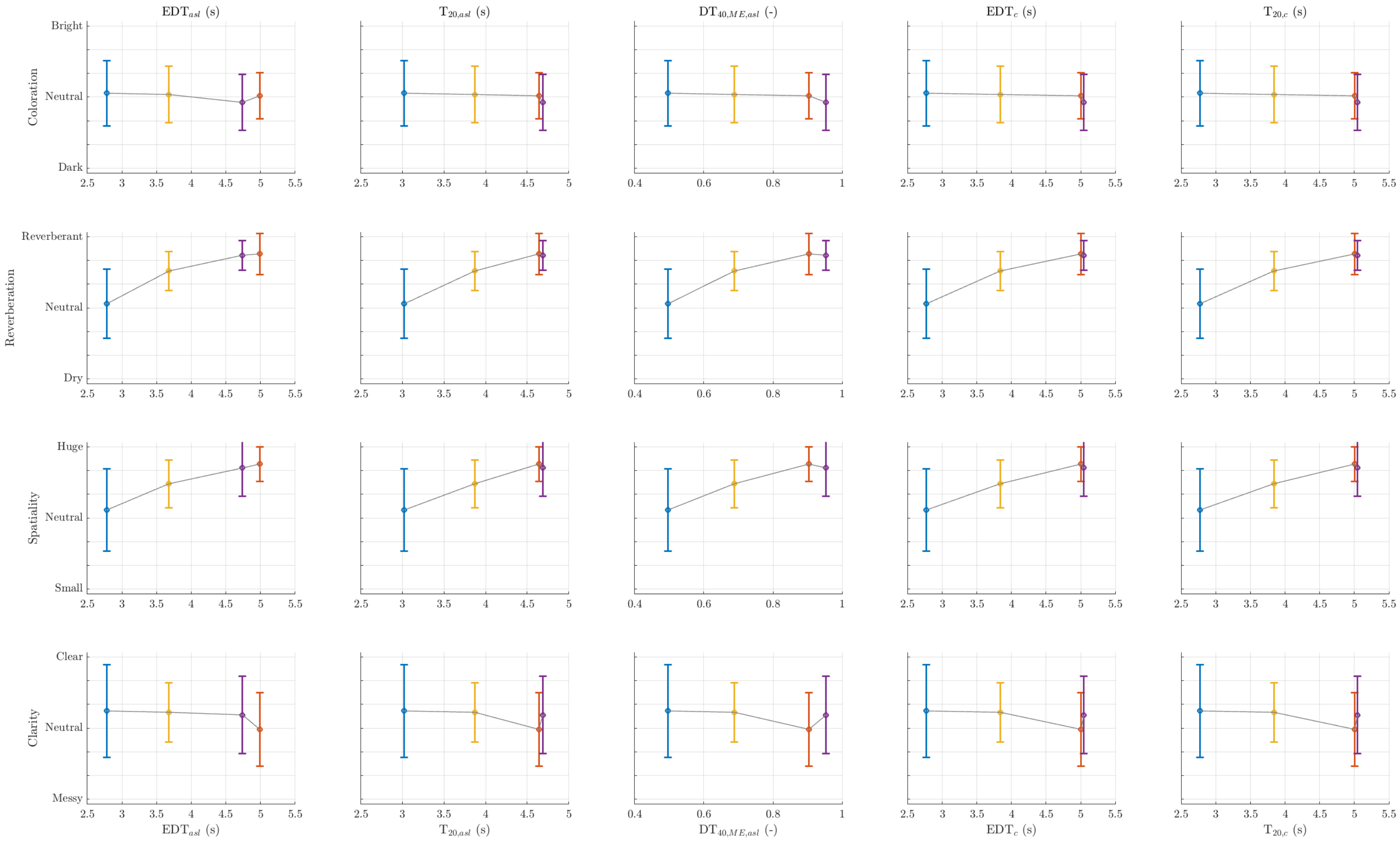


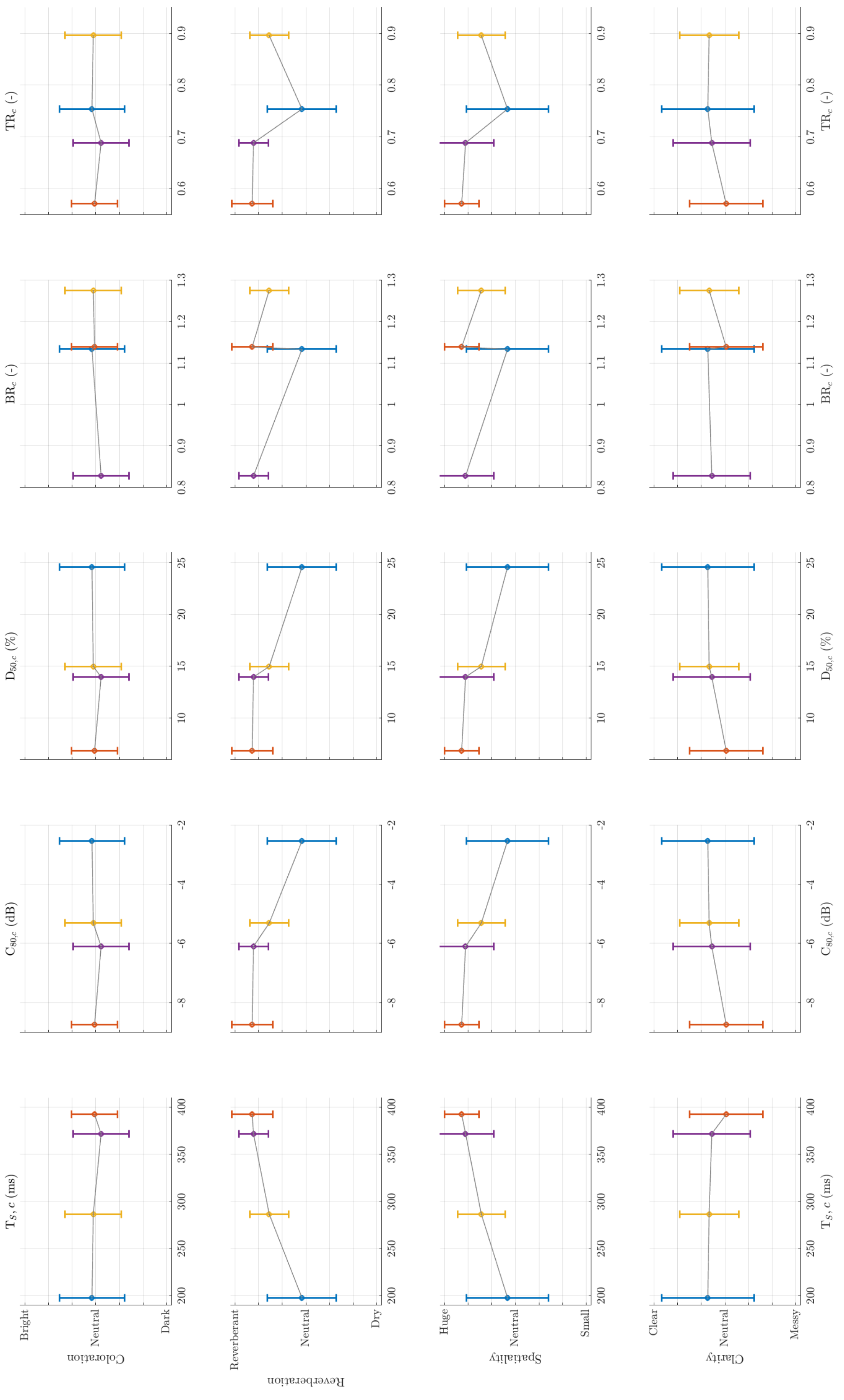
Figure 5.40: Questionnaire in ASL versus Acoustical Parameters (1/6)



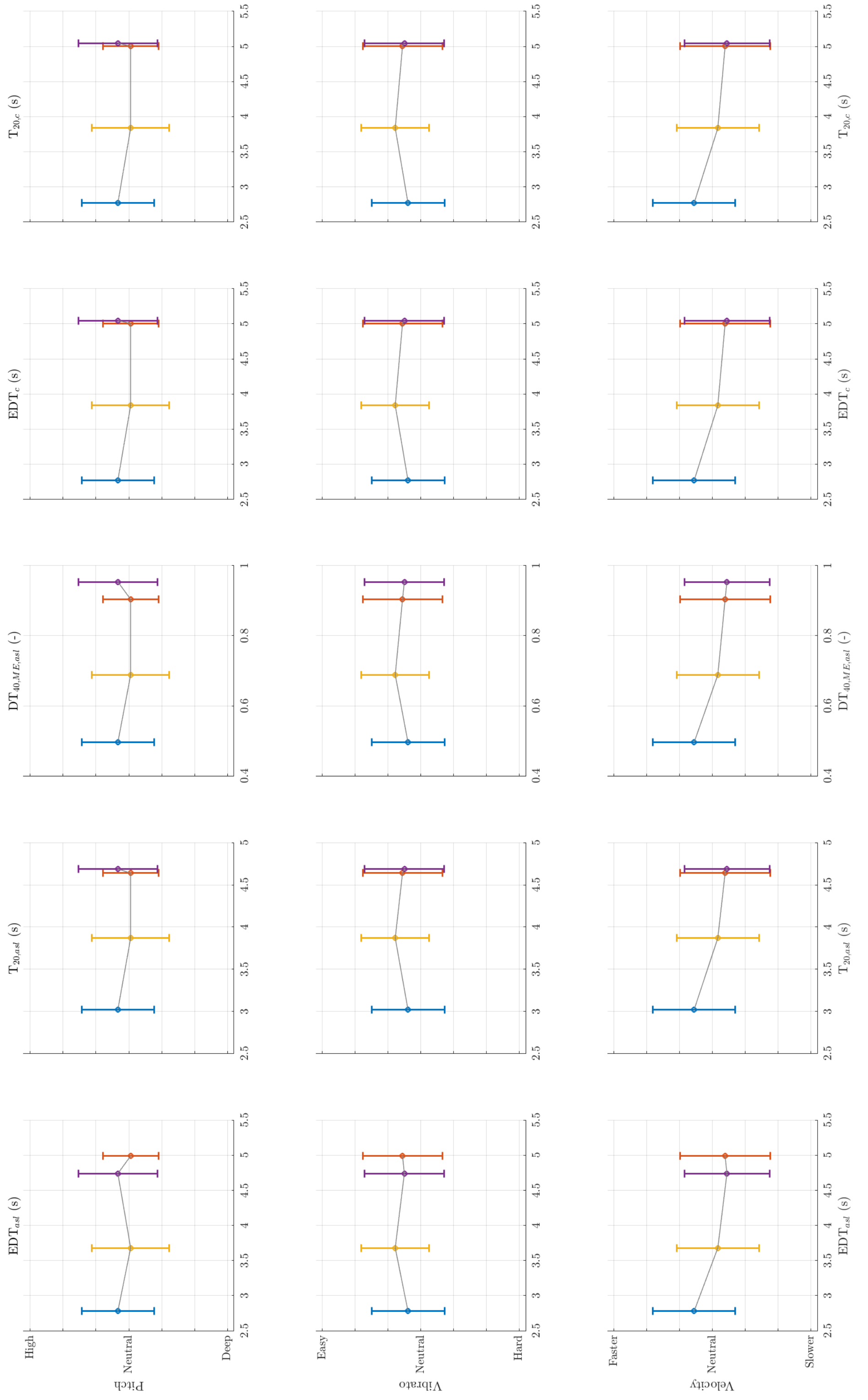
Questionnaire in ASL versus Acoustical Parameters (2/6)



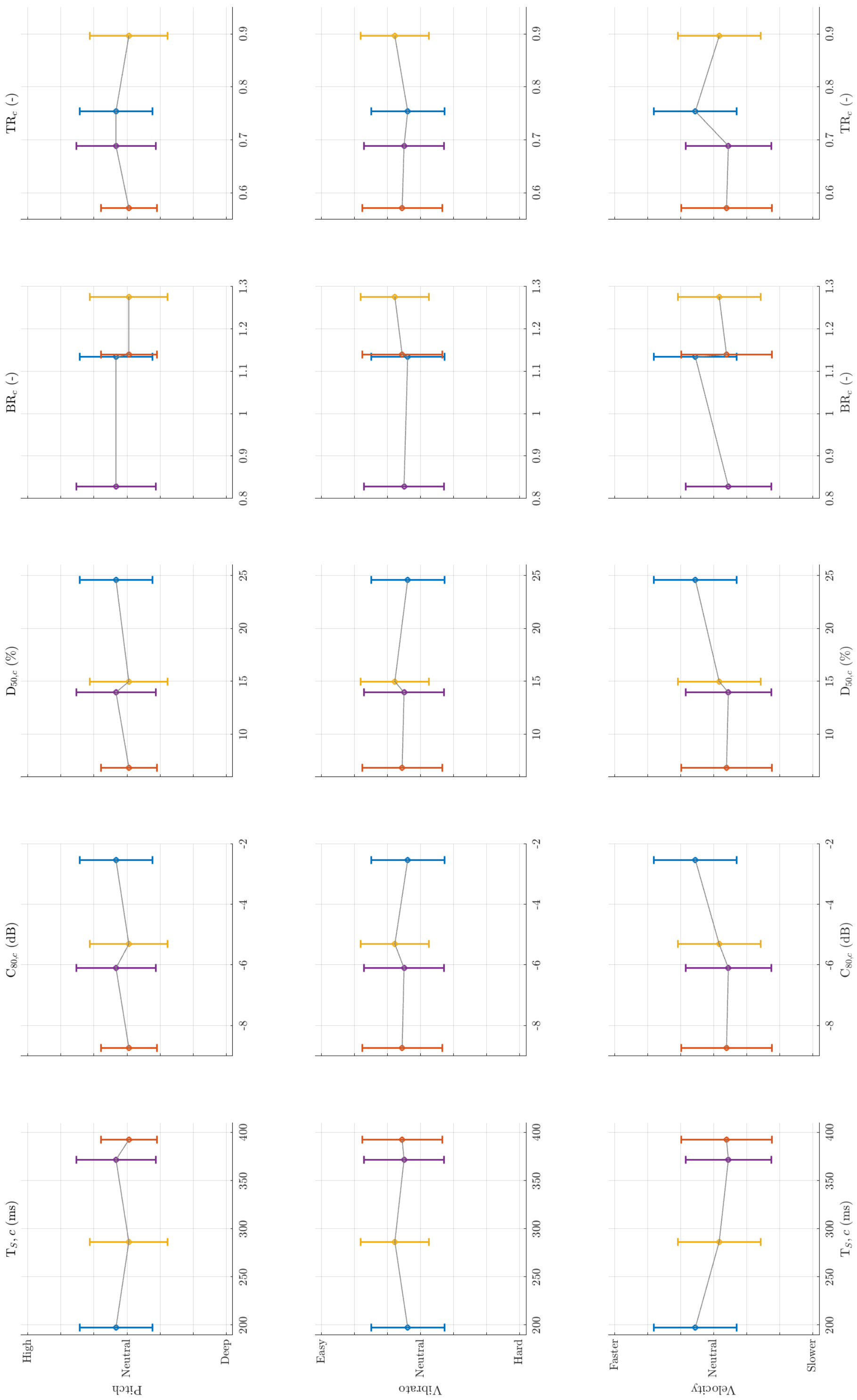
Questionnaire in ASL versus Acoustical Parameters (3/6)



Questionnaire in ASL versus Acoustical Parameters (4/6)



Questionnaire in ASL versus Acoustical Parameters (5/6)



Questionnaire in ASL versus Acoustical Parameters (6/6)

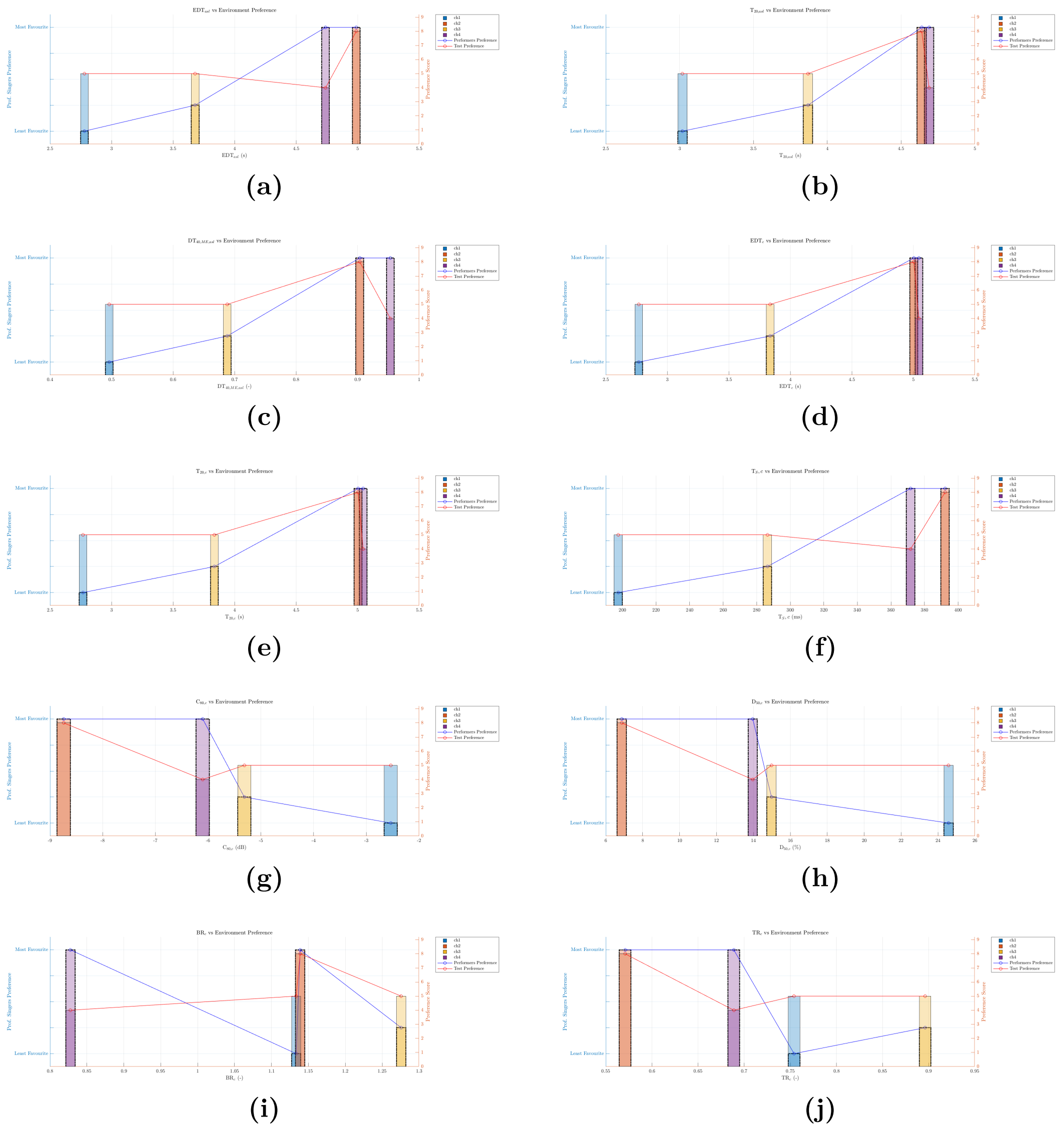
5.7.4 Acoustics versus Singers Preferences

The preferences expressed by both the professional singers in the real environments and the amateur singers who took part in the subjective investigation have been correlated to the acoustical parameters. These graphs are presented in **Figure 5.41**.

For the professional singers *(blue line)* an ascending trend is observable in the graphs of EDT$_{asl}$, T$_{20,asl}$, DT$_{40,ME,asl}$, EDT$_c$, T$_{20,c}$ and T$_{s,c}$, while a descending trend may be observed in the graphs that report the preference in comparison with C$_{80,c}$ and D$_{50,c}$. This findings may be interpreted as a clear preference expressed by the professional singers for the more reverberant environments, which were Church 2 and Church 4.

For what concerns the amateur singers *(red line)*, a similar trend may be observed: the preference they expressed tend to increase accordingly with the Reverbertion Time. It is worth noticing that, from the responses collected, the 4th auralized environment resulted to be disliked compared to the others, in spite of the expected trend.

As in previous sections, BR$_c$ and TR$_c$ do not show significant impact on the preferences expressed by the two groups.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

Figure 5.41: Environment Preference Versus Acoustical Parameters: Professionals and Amateurs

5.7.5 Acoustics versus Soloist Voice Parameters in ASL

The parameters obtained from the testing procedure with the professional soloist singer in the ASL were analyzed in relation to the room acoustics parameters to identify potential correlations.

In **Figure 5.42**, the voice parameters are presented as dependent variables of the room acoustics parameters. While the length of the pauses exhibits notable differences among the four environments, no discernible trends are evident. All other parameters, including the TD, which demonstrated a significant trend when correlated with reverberation objective parameters for the amateur singers, do not exhibit substantial variations across the environments.

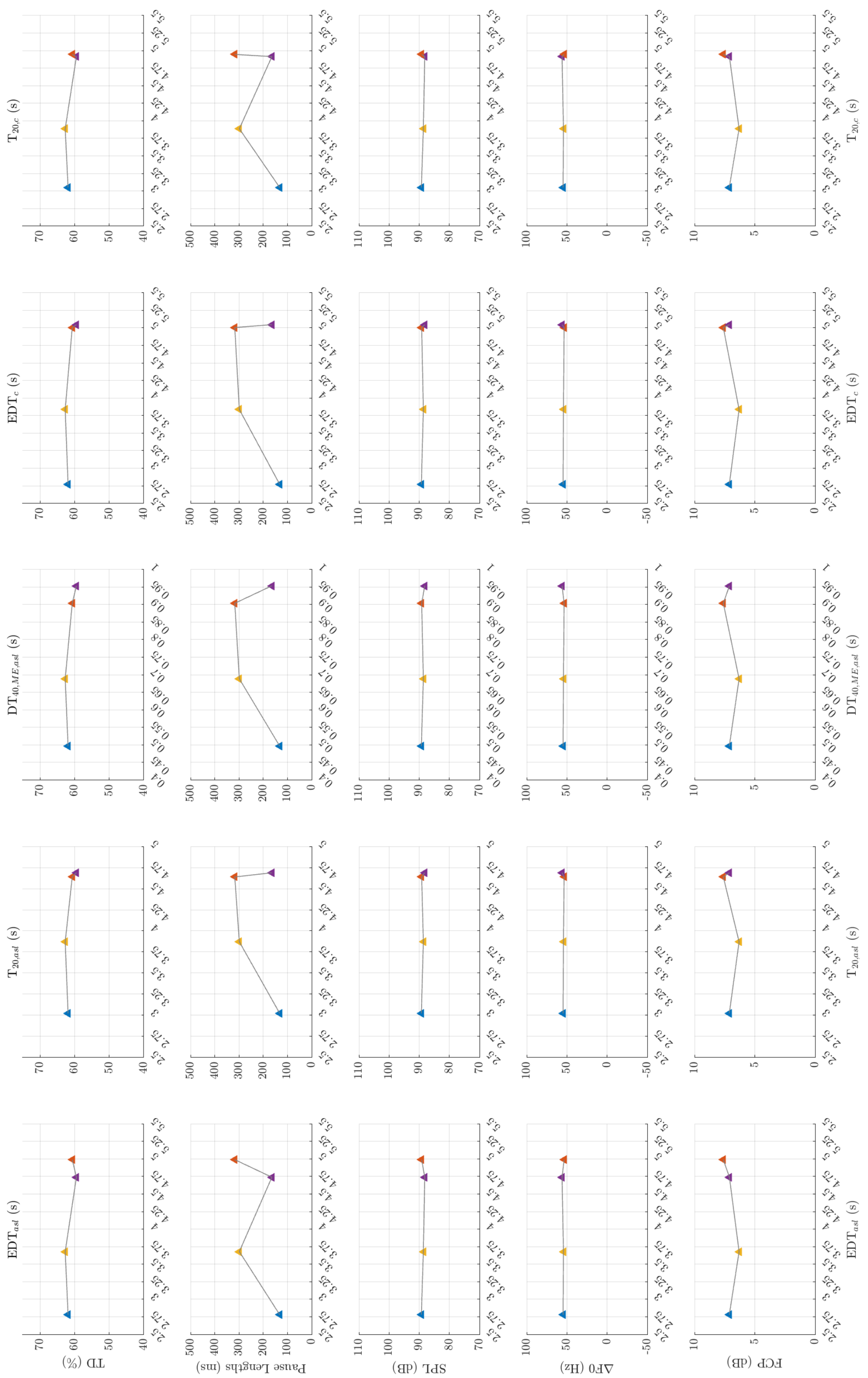
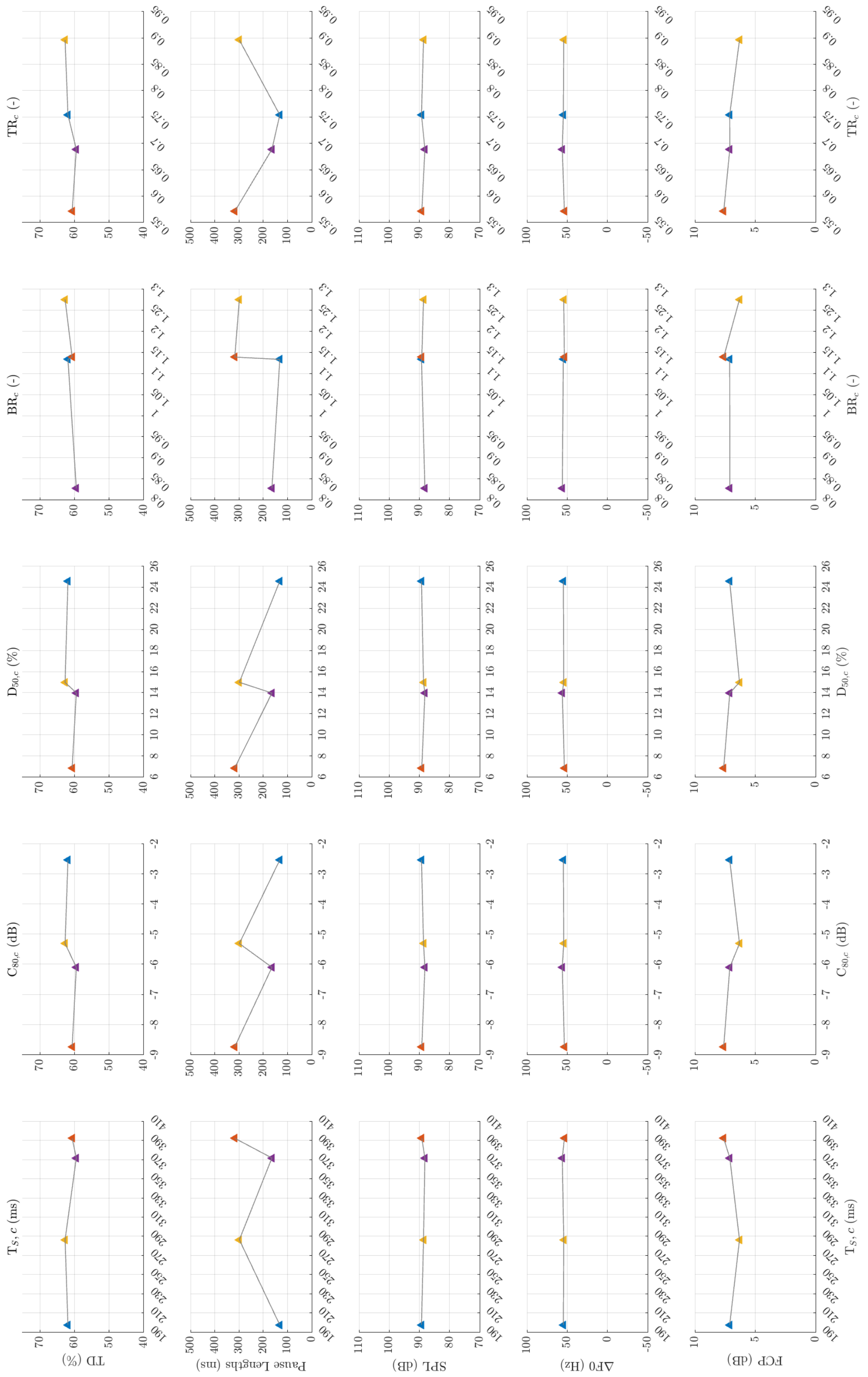
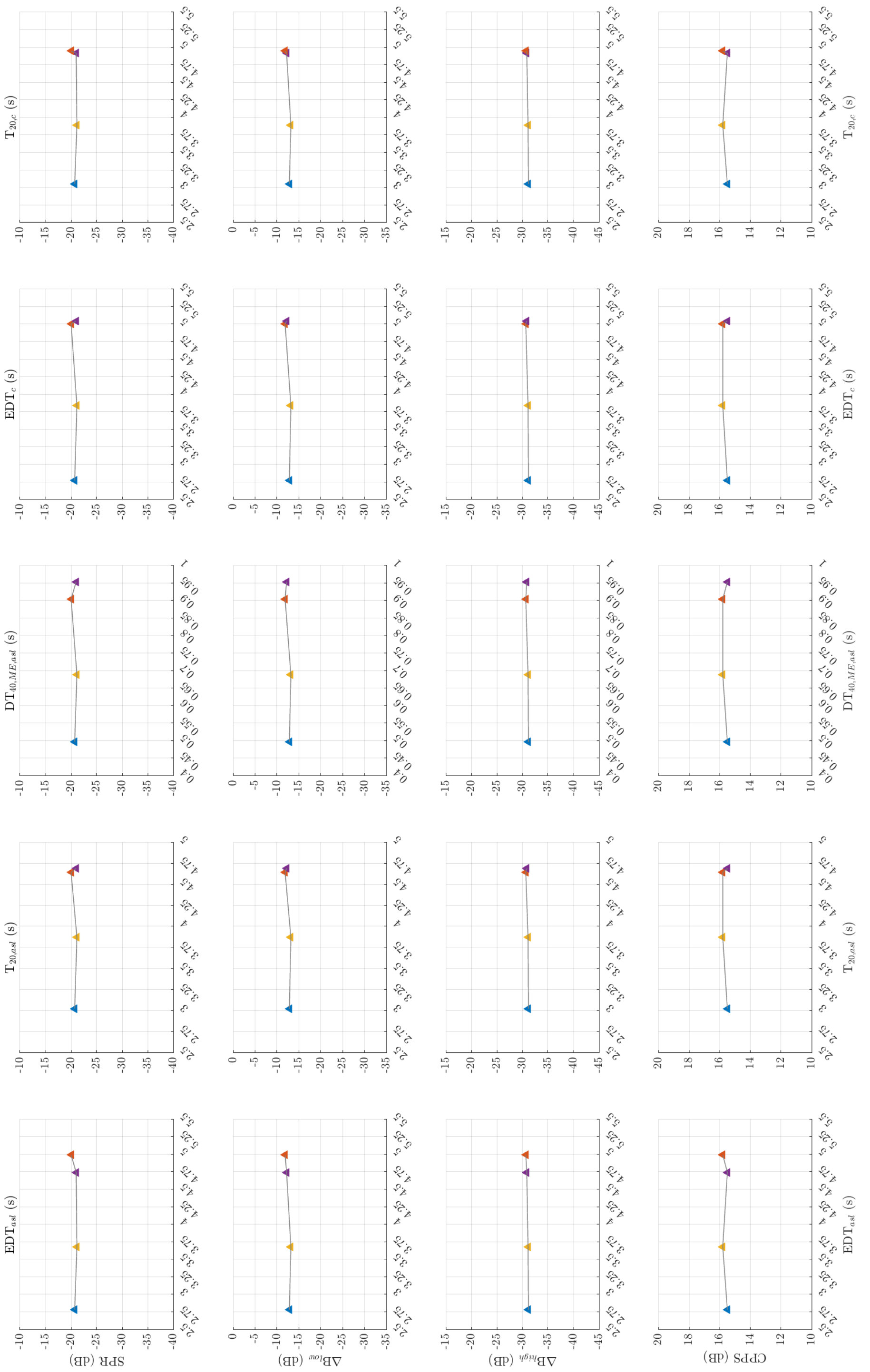


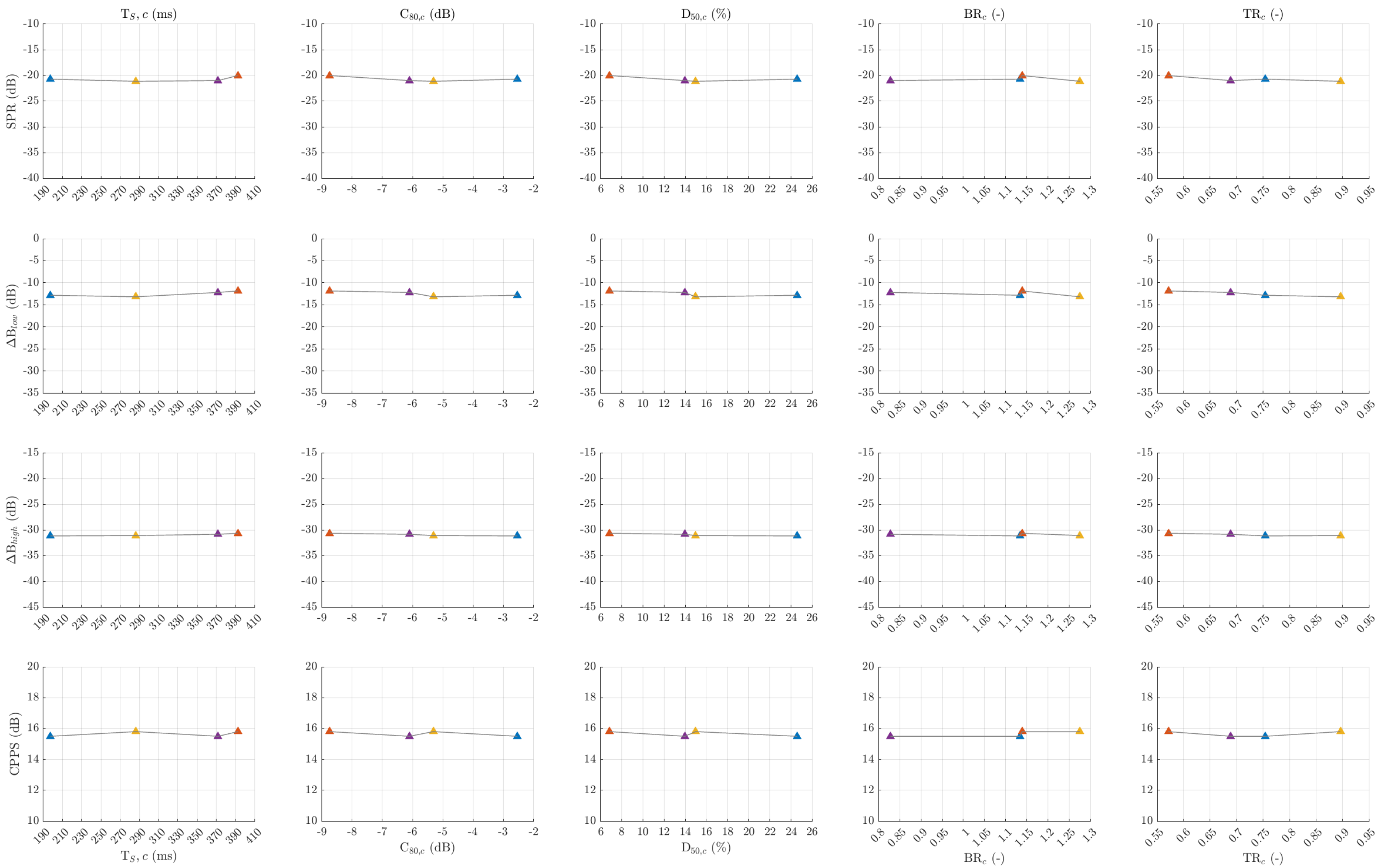
Figure 5.42: Acoustical Parameters compared to Soloist voice parameters in ASL (1/4)



Acoustical Parameters compared to Soloist voice parameters in ASL (2/4)



Acoustical Parameters compared to Soloist voice parameters in ASL (3/4)



Acoustical Parameters compared to Soloist voice parameters in ASL (4/4)

Additionally, the responses to the subjective perceptual questionnaire provided by the Soloist are presented in **Figure 5.43**.

The perceived Three-Dimensionality, which exhibited a distinct ascending trend when compared to reverberation objective parameters in previous analyses, does not display a clear pattern in relation to the same parameters in this context. A similar observation applies to perceived Presence, which shows no significant differences across the four environments, except for Church 2, which presented a lower level.

The perceived Reverberation increases with the Reverberation Time as measured through EDT_{asl}, $\mathrm{T}_{20,asl}$, $\mathrm{DT}_{40,ME,asl}$, EDT_{c}, and $\mathrm{T}_{20,c}$, as well as with the values of $\mathrm{T}_{s,c}$. These findings align with the trends observed among the amateur singers and provide further perceptual validation of the auralization setup. Conversely, perceived Spatiality exhibits a similar trend, with an outlier result for Church 2.

Clarity demonstrates a negative correlation with the objective reverberation measurements, although no clear trend is observed when considering clarity-specific objective parameters such as $\mathrm{C}_{80,c}$ and $\mathrm{D}_{50,c}$.

Lastly, a trend is evident in the Vibrato ratings. This perceptual parameter, intended to reflect the ease of vibration, shows a negative correlation with $\mathrm{T}_{20,asl}$, $\mathrm{DT}_{40,ME,asl}$, EDT_{c}, and $\mathrm{T}_{20,c}$.

Overall, these findings partially corroborate the observations derived from the analysis of the objective parameters as perceived by the amateur singers during the same testing protocol.

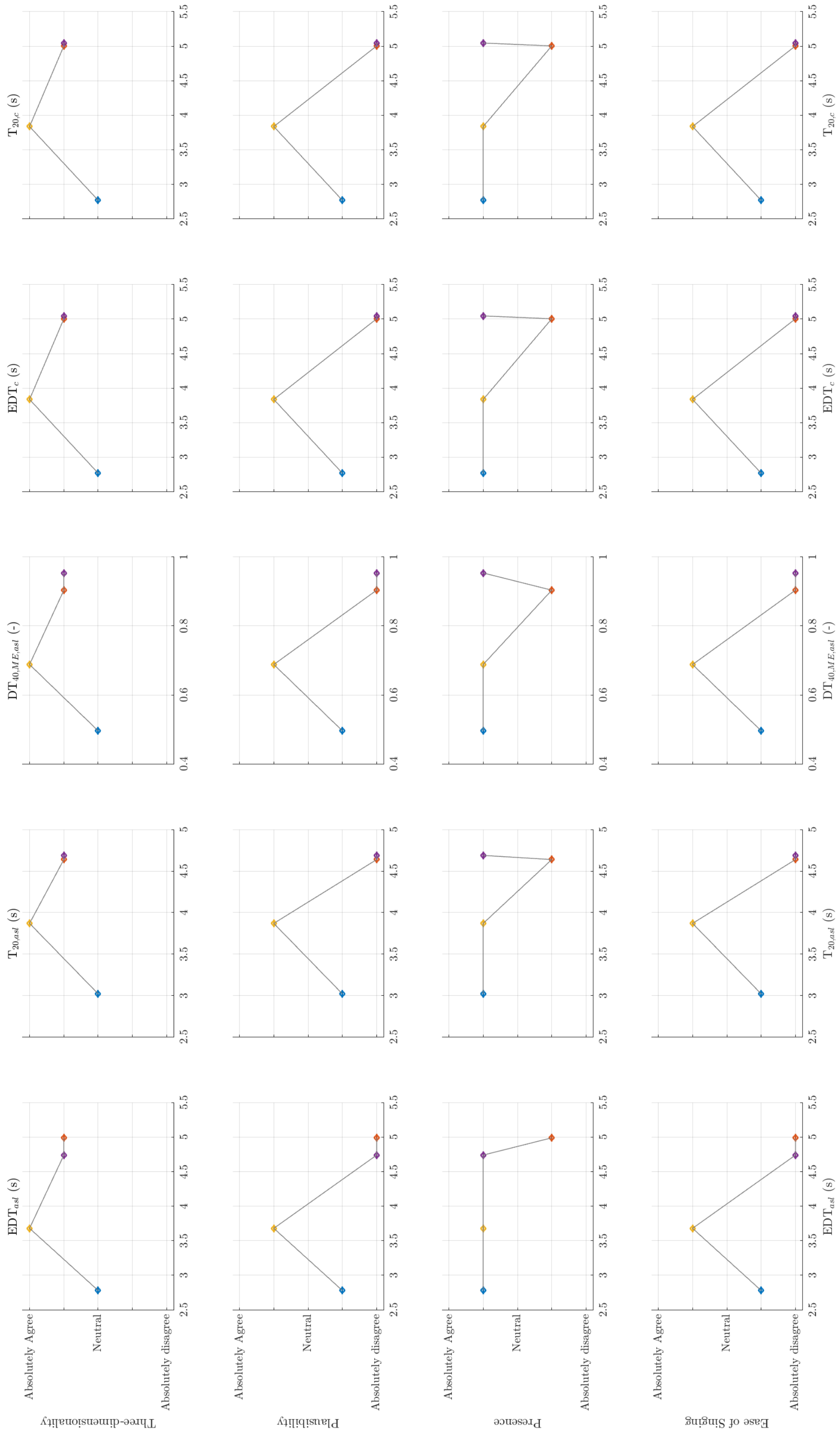
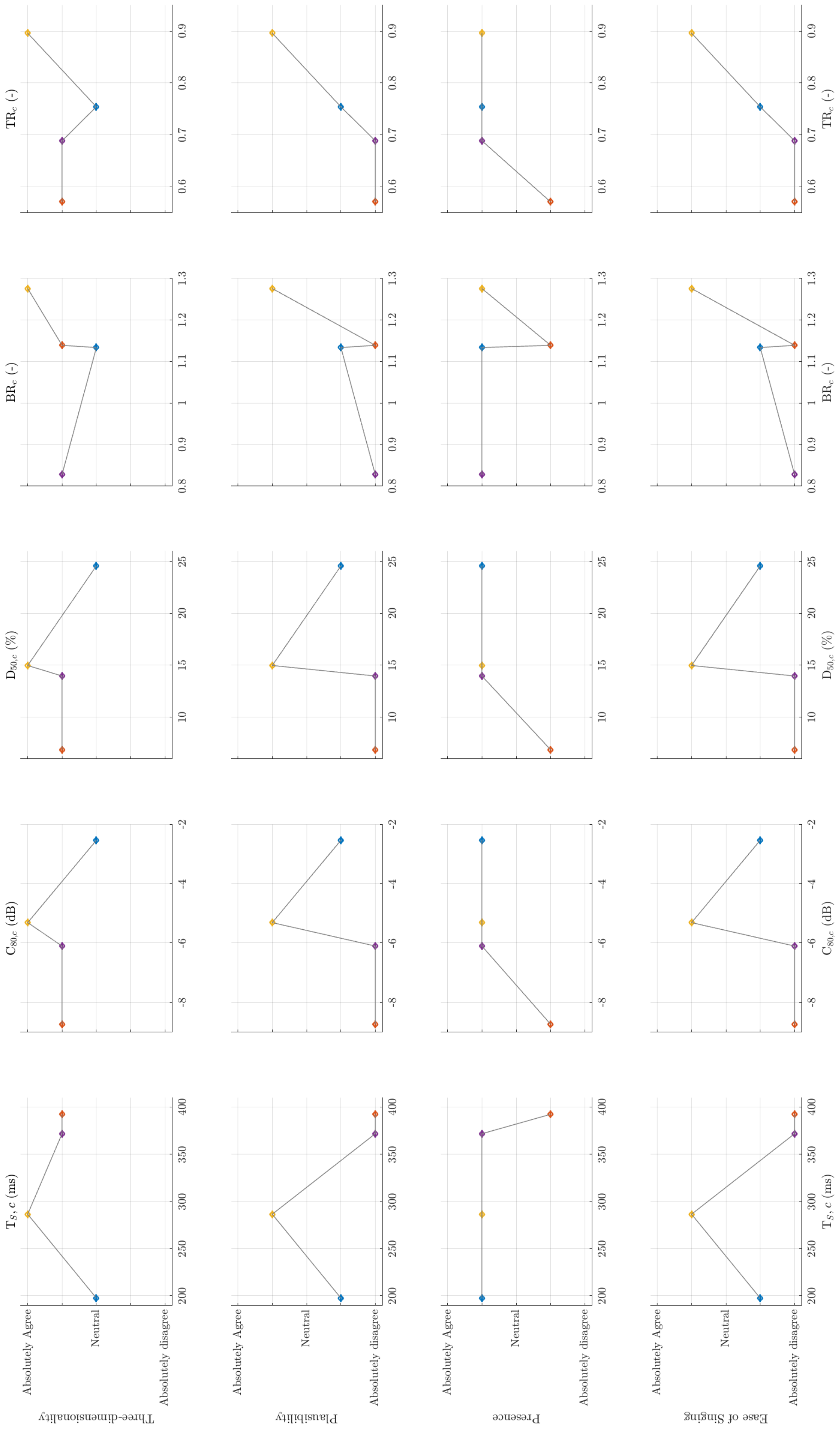
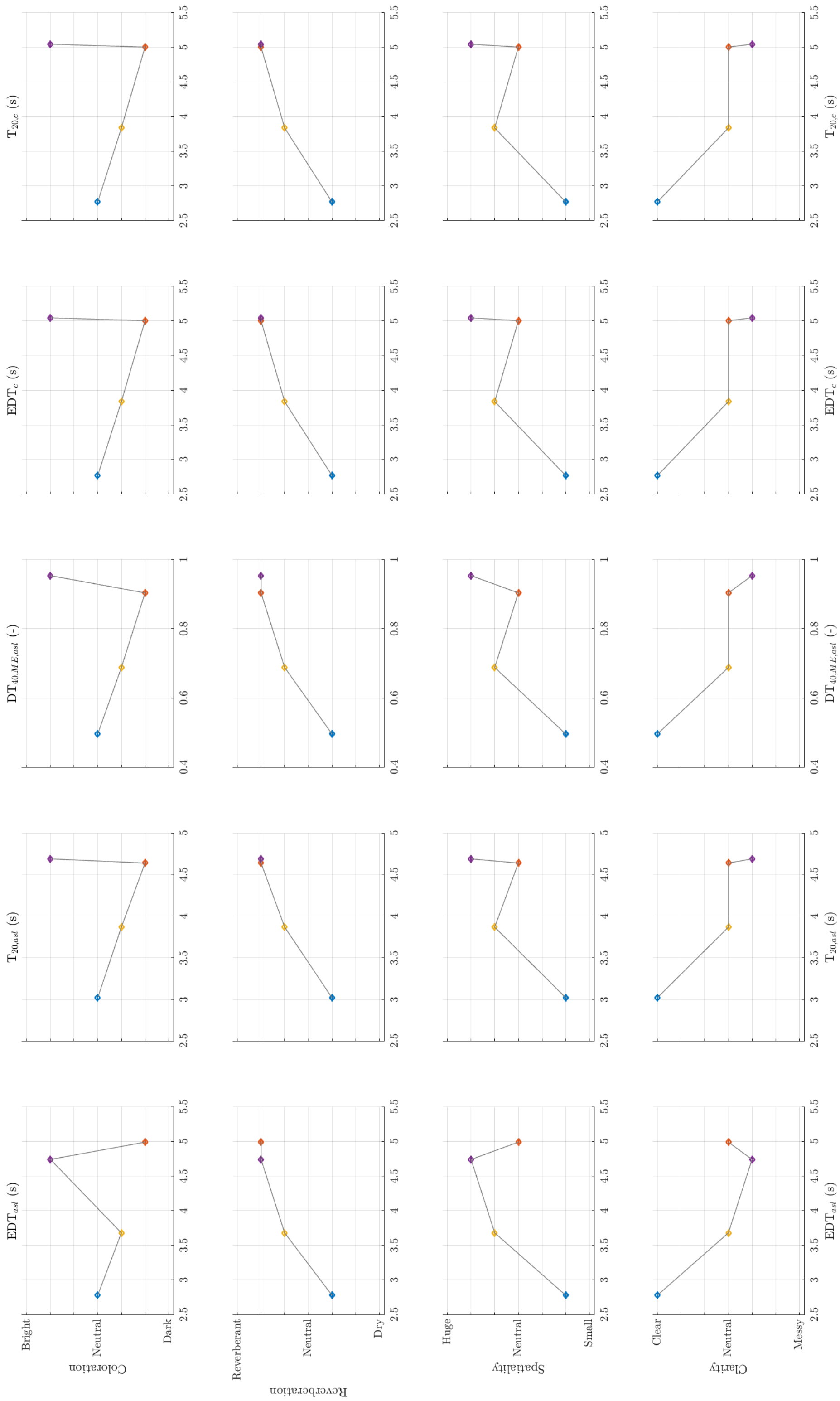


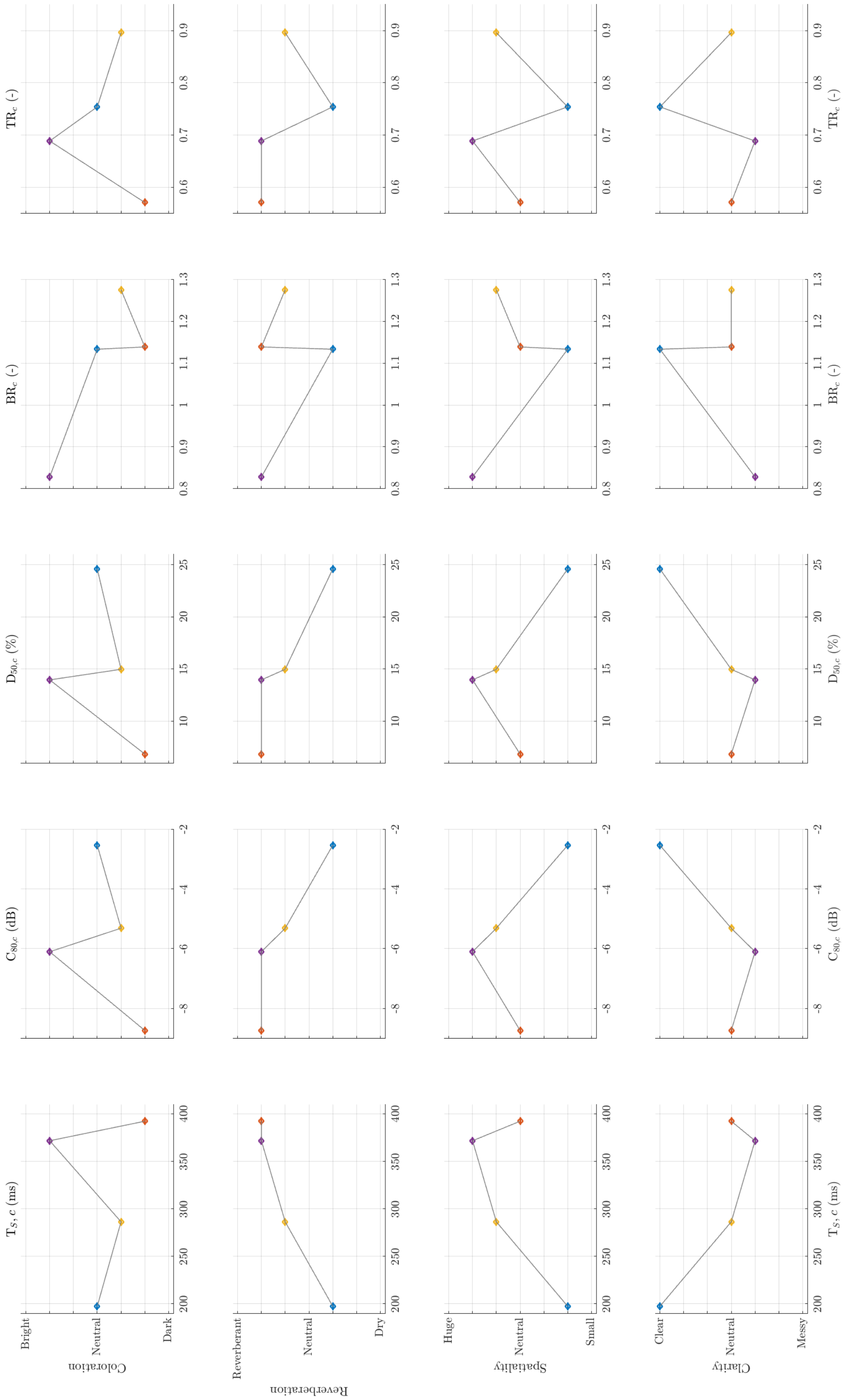
Figure 5.43: Questionnaire of Soloisti n ASL versus Acoustical Parameters (1/6)



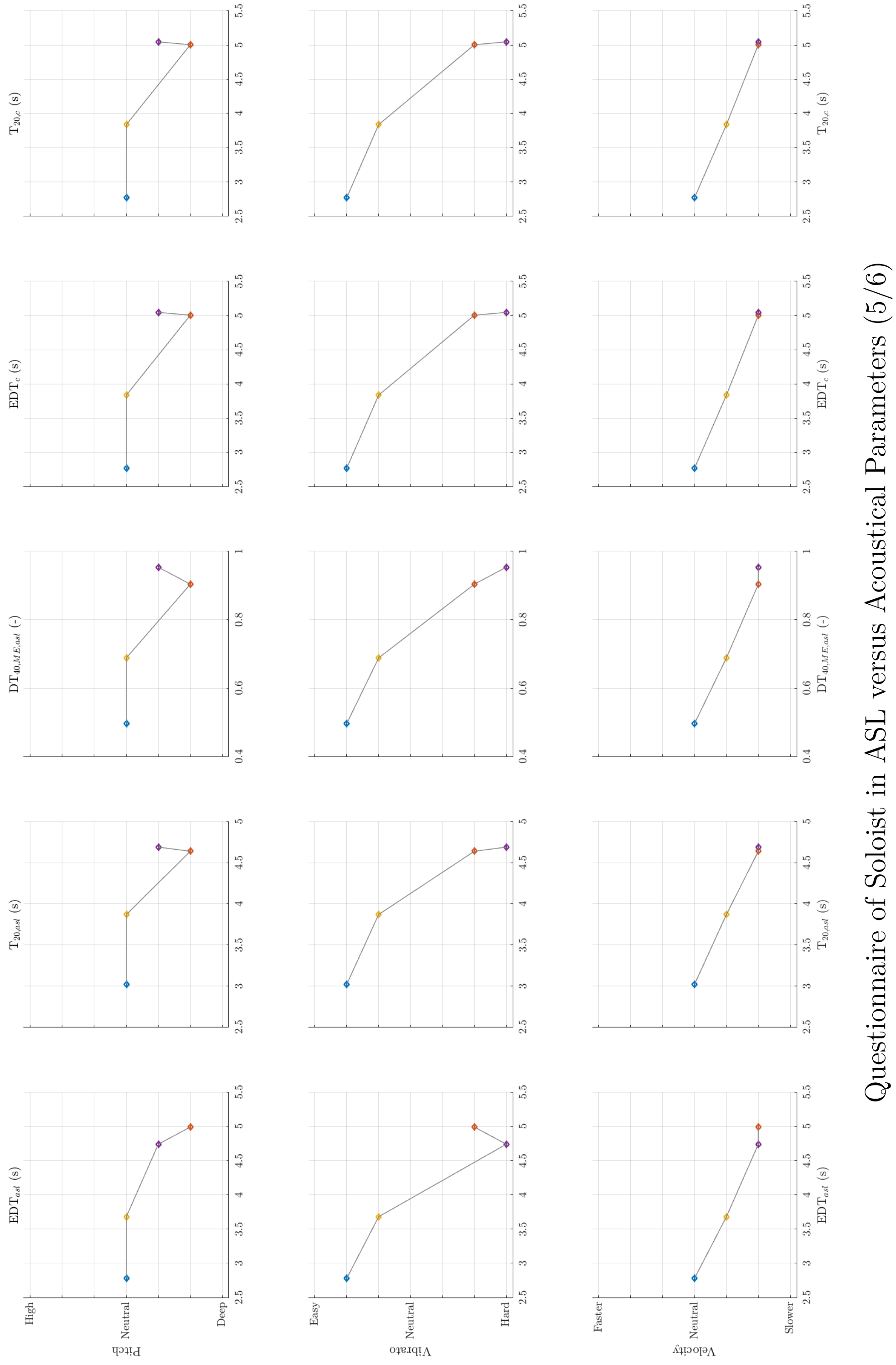
Questionnaire of Soloist in ASL versus Acoustical Parameters (2/6)

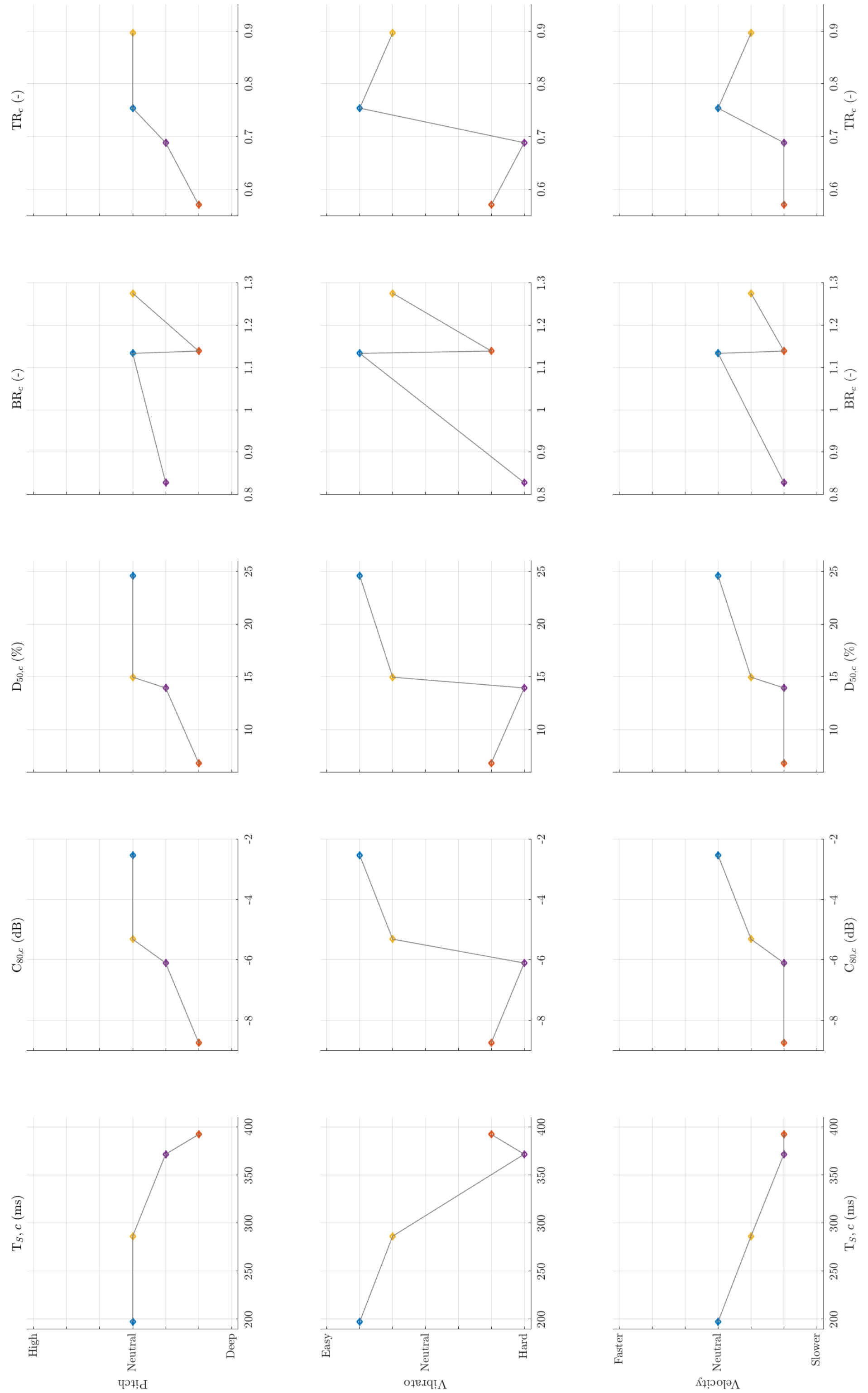


Questionnaire of Soloist in ASL versus Acoustical Parameters (3/6)



Questionnaire of Soloist in ASL versus Acoustical Parameters (4/6)


Questionnaire of Soloist in ASL versus Acoustical Parameters (5/6)
Pitch
High
Neutral
Deep
Vibrato
Easy
Neutral
Hard
Velocity
Faster
Neutral
Slower
EDT$_{asl}$ (s)
T$_{20,asl}$ (s)
DT$_{40,ME,asl}$ (-)
EDT$_{c}$ (s)
T$_{20,c}$ (s)
2.5
3
3.5
4
4.5
5
5.5
0.4
0.6
0.8
1



Questionnaire of Soloist in ASL versus Acoustical Parameters (6/6)

Chapter 6

Future Improvements

This thesis presented the design and implementation process of a Real-Time Auralization System in the 3OA loudspeaker array of the ASL. As this was an exploratory study, certain compromises were made to simplify some of the procedures. Future improvements could enhance the realism of the simulations, aiming for an increased level of fidelity and progressing toward a true reconstruction of actual environments. In this chapter, the author would like to suggest potential directions for future research, based on the findings of this study and hints from past literature.

Yadav et al. [4] auralized the human voice through a binaural output system. Their experimental setup differed significantly from the one used in the ASL. They employed a dummy head, which was progressively rotated to create a map of binaural IRs. These were then used for real-time convolution, with a head-tracking system dynamically selecting the appropriate IR based on head rotation.

The scanning method using a dummy head could be adapted to generate a 3OA IR by combining multiple binaural IRs recorded at different ear positions. This approach could achieve an accurate capture of the IR of a space excited by a (simulated) speaker emission. However, it would be crucial to account for the filtering effects introduced by the dummy head's recording apparatus. Additionally, the role of direct sound propagation via internal conduction within the head itself warrants further investigation.

Kato et al. [20] employed a less precise sampling method, using a directional microphone oriented in six orthogonal directions. However, the idea of utilizing a directional microphone instead of a dummy head recording system is worth consideration. The polar pattern and filtering effects introduced by the measurement microphone should be carefully accounted for during post-processing. Nevertheless, a dedicated framework would be required to obtain the 3OA IRs necessary for the auralization process in this reproduction system.

Another potential area for improvement is the consideration of the human voice's directional characteristics. In this study, the singers' voices were convolved with the sampled IRs, effectively assuming that the human vocal apparatus behaves as an omnidirectional source that excites the environment equally in all 16 directions. *Postma et al.* [8], for example, accounted for source directivity in their auralized listening tests. In our experimental procedure, this simplification did not appear to negatively impact the results, likely due to the measurement methodology used to record the IRs. Participants in the subjective evaluation did not report unrealistic feedback. On the contrary, some singers noted that the perceived sound from behind their ears was lower than expected.

Finally, the author wishes to highlight a significant challenge in achieving realistic auralization: the latency introduced by the auralization engine. In particular, when using loudspeaker arrays, latency throughout the processing chain can make it difficult to faithfully reproduce early reflections from nearby surfaces. For example, a reflection from the ground should reach the performer's ears in less than 6 ms, assuming a mouth height of 1 m and a mean speed of sound of 343 m/s. With advancements in Information Technology, this limitation may be mitigated by future generations of general-purpose computing devices. However, the development of specialized audio hardware optimized for minimizing multi-channel input/output latency could represent a major breakthrough in the pursuit of high-quality, realistic auralization through loudspeaker arrays.

Chapter 7

Conclusions

This thesis aimed to verify the feasibility of an auralization engine in the Audio Space Lab (ASL) of Politecnico di Torino, specifically for applications in human singing auralization. Third-order Ambisonics impulse responses from four actual churches were modified and used in a calibrated real-time convolution system, and subjective singing tests were conducted to assess the resulting auralization quality. Measurements of Reverberation Time in the auralized environments showed that differences between the values from the actual churches and those from the auralized environments were found to be less than 5%.

Eighteen amateur singers participated in a subjective singing test in the ASL. They performed and rated their perception of the feedback provided by the Laboratory's loudspeaker array, yielding encouraging results. Their recordings were collected, along with recordings of the soloist performing in both the actual churches and the laboratory, and a set of vocal features was extracted.

Significant trends in perceived Reverberation and Spatiality were observed among users of the auralization engine, both amateur and professional singers performing in the ASL. However, while professionals reported a perceived ease in vibrating that was negatively correlated with objective reverberation parameters, amateurs noticed a positive influence of this same group of parameters on the sensation of Three-Dimensionality and Presence, understood as an index of perceived immersion in the virtual acoustic environment.

A secondary aim of this study was to investigate singers' adaptation to different venue acoustics, with vocal parameters hypothesized as potential indicators of these processes. Correlations between vocal parameters—considered as indicators of singing quality—yielded few significant results: Time Dose, expressed as the percentage of vocal emission during the duration of a musical piece, was the only parameter to show a slight correlation among amateur singers, but no analogous significant behavior was found in the recordings of the professional soloist performing in the ASL.

The participants' preference rankings closely aligned with those expressed by professional singers who had experienced the actual sampled environments.

These findings leave room for further investigation into the factors influencing singers' perception of performance venues and how performers adjust their vocal features to accommodate different acoustic characteristics.

Appendix A

Subjective Investigation in ASL: questionnaire

In this section, the questionnaire used in the experimental protocol in the ASL is reported.

Politecnico di Torino, DENERG **Febbraio, 2025**

CANTO IN AMBIENTE AURALIZZATO – QUESTIONARIO SOGGETTIVO

ID:…………………

Istruzioni

Viene richiesto di:

1) Indicare sul modulo, in una scala 1-5 o 1-7, **le caratteristiche di ogni ambiente**
2) Indicare sul modulo **che tipo di somiglianze o differenze** sono state percepite tra i diversi ambienti
3) Indicare sul modulo quale/i **ambiente/i si preferiscono**

AMBIENTE 1:

D1.1: Mi è sembrato di percepire un suono tridimensionale:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?..

D1.2: La tua esperienza uditiva in laboratorio sembrava nel complesso verosimile, rispetto a tue esperienze di canto in ambienti reali:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?..

D1.3: Mi sono sentito presente nello spazio:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?..

D1.4: Ho trovato poco faticoso cantare in questo ambiente:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?..

Commenti liberi riguardo all’ambiente:

D1.5: Come valuti le seguenti caratteristiche nell'ambiente?

Definizione dei parametri:

- **Colorazione:** Alterazione del timbro sonoro dovuta a effetti di filtraggio, risonanze o distorsioni nel sistema di riproduzione o nell'ambiente.
  es. scuro: cassa da discoteca, enfatizza i suoni gravi
  metà: ascolto normale
  chiaro: cassa di bassa qualità, stridula, enfatizza i suoni acuti
- **Spazialità:** La percezione della dimensione geometrica dello spazio sonoro che ti circonda.
  es. piccolo: stanza piccola
  medio: stanza media
  grande: stanza molto ampia
- **Riverberazione:** Il prolungamento del suono dovuto alla riflessione sulle superfici dell'ambiente, che ne influenza la percezione di grandezza e profondità.
  es. assorbente: studio di registrazione
  metà: stanza normale
  riverberante: cattedrale
- **Nitidezza:** la possibilità di percepire nitidamente note suonate in successione rapida, influenzata dalla presenza di rumori, distorsioni o dalla qualità della sorgente audio.
  es. confuso: i suoni non sono chiaramente distinguibili, c'è del rumore
  metà: i suoni sono comprensibili ma c'è un effetto di disturbo
  pulito: i suoni sono comprensibili e senza disturbi
- **Pitch:** L'altezza percepita di un suono, determinata dalla sua frequenza fondamentale.
  es. grave: le note sembrano essere generalmente basse
  medio: le note rispecchiano l'originale
  acuto: le note sembrano essere generalmente alte
- **Vibrato:** Una modulazione periodica della frequenza di un suono, spesso usata come tecnica espressiva nella voce e negli strumenti musicali.
  es. difficile: l'ambiente sembra renderti più faticoso vibrare
  metà: l'ambiente non influenza la tua vibrazione
  facile: l'ambiente sembra facilitarti nel vibrare
- **Velocità:** La rapidità con cui un suono o una sequenza di suoni viene riprodotta, influenzando la percezione della durata e del ritmo.
  es. lento: senti di avere modificato il tuo canto, rallentandolo
  veloce: senti di avere modificato il tuo canto, velocizzandolo

Colorazione	**(scuro)** □ □ □ □ □ □ □ **(chiaro)**
Spazialità	**(piccolo)** □ □ □ □ □ □ □ **(grande)**
Riverberazione	**(assorbente)** □ □ □ □ □ □ □ **(riverberante)**
Nitidezza	**(confuso)** □ □ □ □ □ □ □ **(pulito, chiaro)**
Pitch	**(grave)** □ □ □ □ □ □ □ **(acuto)**
Vibrato	**(difficile)** □ □ □ □ □ □ □ **(facile)**
Velocità	**(lento)** □ □ □ □ □ □ □ **(veloce)**

AMBIENTE 2:

D2.1: Mi è sembrato di percepire un suono tridimensionale:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D2.2: La tua esperienza uditiva in laboratorio sembrava nel complesso verosimile, rispetto a tue esperienze di canto in ambienti reali:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D2.3: Mi sono sentito presente nello spazio:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D2.4: Ho trovato poco faticoso cantare in questo ambiente:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

Commenti liberi riguardo all’ambiente:

D2.5: Come valuti le seguenti caratteristiche nell'ambiente?

Definizione dei parametri:

- **Colorazione:** Alterazione del timbro sonoro dovuta a effetti di filtraggio, risonanze o distorsioni nel sistema di riproduzione o nell'ambiente.
  es. scuro: cassa da discoteca, enfatizza i suoni gravi
  metà: ascolto normale
  chiaro: cassa di bassa qualità, stridula, enfatizza i suoni acuti
- **Spazialità:** La percezione della dimensione geometrica dello spazio sonoro che ti circonda.
  es. piccolo: stanza piccola
  medio: stanza media
  grande: stanza molto ampia
- **Riverberazione:** Il prolungamento del suono dovuto alla riflessione sulle superfici dell'ambiente, che ne influenza la percezione di grandezza e profondità.
  es. assorbente: studio di registrazione
  metà: stanza normale
  riverberante: cattedrale
- **Nitidezza:** la possibilità di percepire nitidamente note suonate in successione rapida, influenzata dalla presenza di rumori, distorsioni o dalla qualità della sorgente audio.
  es. confuso: i suoni non sono chiaramente distinguibili, c'è del rumore
  metà: i suoni sono comprensibili ma c'è un effetto di disturbo
  pulito: i suoni sono comprensibili e senza disturbi
- **Pitch:** L'altezza percepita di un suono, determinata dalla sua frequenza fondamentale.
  es. grave: le note sembrano essere generalmente basse
  medio: le note rispecchiano l'originale
  acuto: le note sembrano essere generalmente alte
- **Vibrato:** Una modulazione periodica della frequenza di un suono, spesso usata come tecnica espressiva nella voce e negli strumenti musicali.
  es. difficile: l'ambiente sembra renderti più faticoso vibrare
  metà: l'ambiente non influenza la tua vibrazione
  facile: l'ambiente sembra facilitarti nel vibrare
- **Velocità:** La rapidità con cui un suono o una sequenza di suoni viene riprodotta, influenzando la percezione della durata e del ritmo.
  es. lento: senti di avere modificato il tuo canto, rallentandolo
  veloce: senti di avere modificato il tuo canto, velocizzandolo

Colorazione	**(scuro)** □ □ □ □ □ □ □ **(chiaro)**
Spazialità	**(piccolo)** □ □ □ □ □ □ □ **(grande)**
Riverberazione	**(assorbente)** □ □ □ □ □ □ □ **(riverberante)**
Nitidezza	**(confuso)** □ □ □ □ □ □ □ **(pulito, chiaro)**
Pitch	**(grave)** □ □ □ □ □ □ □ **(acuto)**
Vibrato	**(difficile)** □ □ □ □ □ □ □ **(facile)**
Velocità	**(lento)** □ □ □ □ □ □ □ **(veloce)**

AMBIENTE 3:

D3.1: Mi è sembrato di percepire un suono tridimensionale:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D3.2: La tua esperienza uditiva in laboratorio sembrava nel complesso verosimile, rispetto a tue esperienze di canto in ambienti reali:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D3.3: Mi sono sentito presente nello spazio:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D3.4: Ho trovato poco faticoso cantare in questo ambiente:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

Commenti liberi riguardo all’ambiente:

D3.5: Come valuti le seguenti caratteristiche nell'ambiente?

Definizione dei parametri:

- **Colorazione:** Alterazione del timbro sonoro dovuta a effetti di filtraggio, risonanze o distorsioni nel sistema di riproduzione o nell'ambiente.
  es. scuro: cassa da discoteca, enfatizza i suoni gravi
  metà: ascolto normale
  chiaro: cassa di bassa qualità, stridula, enfatizza i suoni acuti
- **Spazialità:** La percezione della dimensione geometrica dello spazio sonoro che ti circonda.
  es. piccolo: stanza piccola
  medio: stanza media
  grande: stanza molto ampia
- **Riverberazione:** Il prolungamento del suono dovuto alla riflessione sulle superfici dell'ambiente, che ne influenza la percezione di grandezza e profondità.
  es. assorbente: studio di registrazione
  metà: stanza normale
  riverberante: cattedrale
- **Nitidezza:** la possibilità di percepire nitidamente note suonate in successione rapida, influenzata dalla presenza di rumori, distorsioni o dalla qualità della sorgente audio.
  es. confuso: i suoni non sono chiaramente distinguibili, c'è del rumore
  metà: i suoni sono comprensibili ma c'è un effetto di disturbo
  pulito: i suoni sono comprensibili e senza disturbi
- **Pitch:** L'altezza percepita di un suono, determinata dalla sua frequenza fondamentale.
  es. grave: le note sembrano essere generalmente basse
  medio: le note rispecchiano l'originale
  acuto: le note sembrano essere generalmente alte
- **Vibrato:** Una modulazione periodica della frequenza di un suono, spesso usata come tecnica espressiva nella voce e negli strumenti musicali.
  es. difficile: l'ambiente sembra renderti più faticoso vibrare
  metà: l'ambiente non influenza la tua vibrazione
  facile: l'ambiente sembra facilitarti nel vibrare
- **Velocità:** La rapidità con cui un suono o una sequenza di suoni viene riprodotta, influenzando la percezione della durata e del ritmo.
  es. lento: senti di avere modificato il tuo canto, rallentandolo
  veloce: senti di avere modificato il tuo canto, velocizzandolo

Colorazione	**(scuro)** □ □ □ □ □ □ □ **(chiaro)**
Spazialità	**(piccolo)** □ □ □ □ □ □ □ **(grande)**
Riverberazione	**(assorbente)** □ □ □ □ □ □ □ **(riverberante)**
Nitidezza	**(confuso)** □ □ □ □ □ □ □ **(pulito, chiaro)**
Pitch	**(grave)** □ □ □ □ □ □ □ **(acuto)**
Vibrato	**(difficile)** □ □ □ □ □ □ □ **(facile)**
Velocità	**(lento)** □ □ □ □ □ □ □ **(veloce)**

AMBIENTE 4:

D4.1: Mi è sembrato di percepire un suono tridimensionale:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D4.2: La tua esperienza uditiva in laboratorio sembrava nel complesso verosimile, rispetto a tue esperienze di canto in ambienti reali:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D4.3: Mi sono sentito presente nello spazio:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

D4.4: Ho trovato <u>poco</u> faticoso cantare in questo ambiente:

(completamente in disaccordo) □ □ □ □ □ **(concordo completamente)**

Commento?...

Commenti liberi riguardo all'ambiente:

D4.5: Come valuti le seguenti caratteristiche nell'ambiente?

Definizione dei parametri:

- **Colorazione:** Alterazione del timbro sonoro dovuta a effetti di filtraggio, risonanze o distorsioni nel sistema di riproduzione o nell'ambiente.
  es. scuro: cassa da discoteca, enfatizza i suoni gravi
  metà: ascolto normale
  chiaro: cassa di bassa qualità, stridula, enfatizza i suoni acuti
- **Spazialità:** La percezione della dimensione geometrica dello spazio sonoro che ti circonda.
  es. piccolo: stanza piccola
  medio: stanza media
  grande: stanza molto ampia
- **Riverberazione:** Il prolungamento del suono dovuto alla riflessione sulle superfici dell'ambiente, che ne influenza la percezione di grandezza e profondità.
  es. assorbente: studio di registrazione
  metà: stanza normale
  riverberante: cattedrale
- **Nitidezza:** la possibilità di percepire nitidamente note suonate in successione rapida, influenzata dalla presenza di rumori, distorsioni o dalla qualità della sorgente audio.
  es. confuso: i suoni non sono chiaramente distinguibili, c'è del rumore
  metà: i suoni sono comprensibili ma c'è un effetto di disturbo
  pulito: i suoni sono comprensibili e senza disturbi
- **Pitch:** L'altezza percepita di un suono, determinata dalla sua frequenza fondamentale.
  es. grave: le note sembrano essere generalmente basse
  medio: le note rispecchiano l'originale
  acuto: le note sembrano essere generalmente alte
- **Vibrato:** Una modulazione periodica della frequenza di un suono, spesso usata come tecnica espressiva nella voce e negli strumenti musicali.
  es. difficile: l'ambiente sembra renderti più faticoso vibrare
  metà: l'ambiente non influenza la tua vibrazione
  facile: l'ambiente sembra facilitarti nel vibrare
- **Velocità:** La rapidità con cui un suono o una sequenza di suoni viene riprodotta, influenzando la percezione della durata e del ritmo.
  es. lento: senti di avere modificato il tuo canto, rallentandolo
  veloce: senti di avere modificato il tuo canto, velocizzandolo

Colorazione	**(scuro)**	□ □ □ □ □ □ □	**(chiaro)**
Spazialità	**(piccolo)**	□ □ □ □ □ □ □	**(grande)**
Riverberazione	**(assorbente)**	□ □ □ □ □ □ □	**(riverberante)**
Nitidezza	**(confuso)**	□ □ □ □ □ □ □	**(pulito, chiaro)**
Pitch	**(grave)**	□ □ □ □ □ □ □	**(acuto)**
Vibrato	**(difficile)**	□ □ □ □ □ □ □	**(facile)**
Velocità	**(lento)**	□ □ □ □ □ □ □	**(veloce)**

Comparare le differenze percepite dopo l’ascolto di tutti e 4 gli ambienti:

S1: Quanto sono diversi, nel complesso, i 4 ambienti?

(nessuna differenza) □ □ □ □ □ **(chiara differenza)**

Quali differenze principali hai percepito?

□ **Colorazione** □ **Spazialità** □ **Riverberazione** □ **Nitidezza**

□ **Pitch** □ **Vibrato** □ **Velocità** □ **altro…………………………….**

S2: C’erano ambienti simili tra loro? Se sì, Quali?

Quali somiglianze hai percepito?

□ **Colorazione** □ **Spazialità** □ **Riverberazione** □ **Nitidezza**

□ **Pitch** □ **Vibrato** □ **Velocità** □ **altro…………………………….**

S3: Indicare

Quale/i ambiente/i preferisci?

..

Perché?...

..

S4: Altri commenti liberi sull’intera esperienza:

Declaration of AI Tools Usage

During the writing of this thesis, the OpenAI ChatGPT artificial intelligence model was utilized solely for textual refinement, including grammar and clarity improvements. The tool was not used to generate ideas, interpretations, or conclusions, all of which remain entirely original and the result of independent research.

Bibliography

[1] Michael Kleiner, Björn-Inge Dalenbäck, and Patrik Svensson. Auralization – an overview. *Journal of the Audio Engineering Society*, 41:861–875, 1993.

[2] Michael Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, Berlin, Germany, 2008.

[3] Michael Vorländer, Dirk Schröder, Simon Pelzer, and Frank Wefers. Virtual reality for architectural acoustics. *Journal of Building Performance Simulation*, 8(1):15–25, 2015.

[4] Manuj Yadav, Densil Cabrera, and William L. Martens. A system for simulating room acoustical environments for one's own voice. *Applied Acoustics*, 73(4):409–414, 4 2012.

[5] Luis A. Miranda Jofre, Densil Cabrera, Manuj Yadav, Anna Sygulska, and William Martens. Evaluation of stage acoustics preference for a singer using oral-binaural room impulse responses. In *Proceedings of the Acoustical Society of America*, pages 015074–015074, Montreal, Canada, 2013.

[6] Barteld N. J. Postma and Brian F. G. Katz. Creation and calibration method of acoustical models for historic virtual reality auralizations. *Virtual Reality*, 19(3–4):161–180, 11 2015.

[7] Barteld N. J. Postma, Andrew Tallon, and Brian F. G. Katz. Calibrated auralization simulation of the abbey of saint-germain-des-prés for historical study. In *Auditorium Acoustics 2015*, Paris, 2015. Institute of Acoustics.

[8] Barteld N. J. Postma and Brian F. G. Katz. Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. *The Journal of the Acoustical Society of America*, 140(6):4326–4337, 12 2016.

[9] S.S. Mullins and B.F.G. Katz. Immersive auralisations for choral ensembles. In *Proceedings of Auditorium Acoustics 2023*, volume 45, 2023.

[10] Brian F. G. Katz, Sandie Leconte, and Peter Stitt. Evaa: A platform for experimental virtual archeological-acoustics to study the influence of performance space. In *International Symposium on Room Acoustics (ISRA)*, Amsterdam, Netherlands, 9 2019.

[11] David Thery and Brian F. G. Katz. Auditory perception stability evaluation

comparing binaural and loudspeaker ambisonic presentations of dynamic virtual concert auralizations. *The Journal of the Acoustical Society of America*, 149(1):246–258, 1 2021.

[12] Nolan Eley, Sarabeth Mullins, Peter Stitt, and Brian F. G. Katz. Virtual notre-dame: Preliminary results of real-time auralization with choir members. In *2021 Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pages 1–6, Bologna, Italy, 2021. IEEE.

[13] Julien De Muynke, Nolan Eley, Julien Ferrando, and Brian F. G. Katz. Preliminary analysis of vocal ensemble performances in real-time historical auralizations of the palais des papes. In *Proceedings of SAAT 2022*, page sat_204, Verona, Italy, 2022. Accessed 21 March 2025.

[14] David Thery, Vincent Boccara, and Brian F. G. Katz. Auralization uses in acoustical design: A survey study of acoustical consultants. *The Journal of the Acoustical Society of America*, 145(6):3446–3456, 6 2019.

[15] Jonas Brunskog, Anders Christian Gade, Gaspar Payá Bellester, and Lilian Reig Calbo. Increase in voice level and speaker comfort in lecture rooms. *The Journal of the Acoustical Society of America*, 125(4):2072–2082, 2009.

[16] David Pelegrín-García and Jonas Brunskog. Speakers' comfort and voice level variation in classrooms: Laboratory research. *The Journal of the Acoustical Society of America*, 132(1):249–260, 2012.

[17] Giuseppina Emma Puglisi, Arianna Astolfi, Lady Catherine Cantor Cutiva, and Alessio Carullo. Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America*, 141(1):441–452, 2017.

[18] S. Ternström. Long-time average spectrum characteristics of different choirs in different rooms. *STL-QPSR*, 30(3):15–31, 1989.

[19] Z. S. Kalkandjiev and S. Weinzierl. The influence of room acoustics on solo music performance: An empirical case study. *Acta Acustica United with Acustica*, 99(3):433–441, 2013.

[20] Kosuke Kato, Kanako Ueno, and Keiji Kawai. Effect of room acoustics on musicians' performance. part ii: Audio analysis of the variations in performed sound signals. *Acta Acustica United with Acustica*, 101(4):743–759, July 2015.

[21] Paul Luizard, Jochen Steffens, and Stefan Weinzierl. Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions? *The Journal of the Acoustical Society of America*, 147(2):EL132–EL137, February 2020.

[22] Paul Luizard and Nathalie Henrich Bernardoni. Changes in the voice production of solo singers across concert halls. *The Journal of the Acoustical Society of America*, 148(1):EL33–39, 2020.

[23] P. Bottalico, N. Łastowiecka, J.D. Glasner, and Y.G. Redman. Singing in

different performance spaces: The effect of room acoustics on vibrato and pitch inaccuracy. *Journal of the Acoustical Society of America*, 151(6):4131–4139, 2022.

[24] Y.G. Redman, J.D. Glasner, D. D'Orazio, and P. Bottalico. Singing in different performance spaces: The effect of room acoustics on singers' perception. *Journal of the Acoustical Society of America*, 154(4):2256–2264, 2023.

[25] Louena Shtrepi, Angela Guastamacchia, Marco Masoero, and Arianna Astolfi. An investigation on the spatial adaptation of an artistic performance in contemporary churches. In *Proceedings of the 2023 International Conference on Immersive and 3D Audio (I3DA)*, pages 1–7. IEEE, 2023.

[26] F. Martellotta, E. Cirillo, A. Carbonari, and P. Ricciardi. Guidelines for acoustical measurements in churches. *Applied Acoustics*, 70:378–388, 2008.

[27] Faber Teater. Faber teater - official website. Accessed: 2025-02-28.

[28] Angelo Farina. A-format to b-format filter matrices - zylia, july 2020, 2020. Accessed: 2025-08-05.

[29] P. Dietrich, M. Guski, J. Klein, M. Müller-Trapet, M. Pollow, R. Scharrer, and M. Vorländer. Measurements and room acoustic analysis with the ita-toolbox for matlab. In *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, page 50, 3 2013.

[30] M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, and M. Vorländer. The ita-toolbox: An open source matlab toolbox for acoustic measurements and signal processing. In *Proceedings of the 43th Annual German Congress on Acoustics*, pages 6–9, Kiel, Germany, 2017.

[31] International Organization for Standardization. ISO 3382-1:2009 - Acoustics — Measurement of room acoustic parameters — Part 1: Performance spaces, 2009. Available at: `https://www.iso.org/standard/40979.html`.

[32] J.S. Bradley. Review of objective room acoustics measures and future needs. *Applied Acoustics*, 72(10):713–720, 10 2011.

[33] Marshall Long. Design of rooms for music. In *Architectural Acoustics*, pages 723–777. Elsevier, 2014.

[34] ODEON. Odeon room acoustics software. Accessed: 2025-02-21.

[35] ODEON. Room acoustics. Accessed: 2025-02-21.

[36] R. Kürer. *Doctoral Dissertation*. PhD thesis, Technical University of Berlin, 1972.

[37] A. Gade. Investigations of musicians' room acoustic conditions in concert halls. pt. ii: Field experiments and synthesis of results. *Acustica*, 69:249–262, 1989.

[38] Tapio Lokki, Jukka Pätynen, Antti Kuusinen, and Sakari Tervo. Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *The Journal of the Acoustical Society of America*, 132(5):3148–3161, 2012.

[39] Angelo Farina. A new audacity feature: Room objective acustical parameters calculation module, 2009. Accessed: 2025-02-28.

[40] A. Guastamacchia, M. Ebri, A. Bottega, E. Armelloni, A. Farina, G. Puglisi, F. Riente, L. Shtrepi, M.C. Masoero, and A. Astolfi. Set up and preliminary validation of a small spatial sound reproduction system for clinical purposes. In *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, pages 4991–4998, Turin, Italy, 2024. European Acoustics Association.

[41] Riccardo Lacqua. Physical acoustical validation of the audio space lab at the polytechnic of turin. Tesi di laurea magistrale, Politecnico di Torino, Turin, Italy, 7 2024. Relatori: Arianna Astolfi, Angela Guastamacchia, Louena Shtrepi.

[42] David Pelegrín-García, Jonas Brunskog, Viveka Lyberg-Åhlander, and Anders Löfqvist. Measurement and prediction of voice support and room gain in school classrooms. *The Journal of the Acoustical Society of America*, 131(1):194–204, 2012.

[43] David Pelegrín-García. Comment on "increase in voice level and speaker comfort in lecture rooms" [j. acoust. soc. am. 125, 2072–2082 (2009)] (l). *The Journal of the Acoustical Society of America*, 129(3):1161–1164, 2011.

[44] A. Nacci, B. Fattori, V. Mancini, E. Panicucci, F. Ursino, F. M. Cartaino, and S. Berrettini. The use and role of the ambulatory phonation monitor (apm) in voice assessment. *Acta Otorhinolaryngologica Italica*, 33(1):49–55, 2013.

[45] Alice Fantoni. Assessment of vocal fatigue of multiple sclerosis patients. validation of a contact microphone-based device for long-term monitoring, 2023. Corso di laurea magistrale in Ingegneria Biomedica. Relatori: Alessio Carullo, Alberto Vallan.

[46] Antonio Romano. (in italian) la qualitÀ della voce. In *VIII Convegno dell'Associazione Italiana Scienze della Voce*, Rome, Italy, 2012.

[47] Ilaria Leocata. Singing voice quality assessment in professional singers through acoustic parameters obtained with different microphones, 2018. Corso di laurea magistrale in Ingegneria Biomedica. Relatori: Alessio Carullo, Alberto Vallan, Antonella Castellana, Arianna Astolfi.

[48] F.M.B. Lã, L.S. Silva, and S. Granqvist. Long-term average spectrum characteristics of portuguese fado-canção from coimbra. *Journal of Voice*, 37(4):631.e7–631.e15, 2023.

[49] Koichi Omori, Ashutosh Kacker, Linda M. Carroll, William D. Riley, and Stanley M. Blaugrund. Singing power ratio: Quantitative evaluation of singing voice quality. *Journal of Voice*, 10(3):228–235, 1 1996.

[50] Christopher Watts, Kathryn Barnes-Burroughs, Julie Estis, and Debra Blanton. The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. *Journal of Voice*, 20(1):82–88, 3

2006.

[51] Antonella Castellana, Alessio Carullo, Simone Corbellini, and Arianna Astolfi. Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel. *IEEE Transactions on Instrumentation and Measurement*, 67(3):646–654, 2018.

[52] Calvin P. Baker, Johan Sundberg, Suzanne C. Purdy, Te Oti Rakena, and Sylvia H. De S. Leão. CPPS and voice-source parameters: Objective analysis of the singing voice. *Journal of Voice*, 38(3):549–560, 5 5.

[53] Plogue. Bidule - the new standard in modular audio software, 2025. Accessed: 2025-02-24.

[54] Angelo Farina. X-mcfx: A tool for multichannel acoustic measurements and processing. Accessed: 2025-02-21.

[55] Adobe Inc. Adobe audition, 2025. Accessed: 2025-02-26.

[56] Shure Inc. Beta 5x user guide, 2025. Accessed: 2025-02-26.