Perttu Laukkanen

Evaluation of Studio Control Room Acoustics with Spatial Impulse Responses and Auralization

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 24.2.2014

Thesis supervisor:

Prof. Tapio Lokki

Thesis advisor:

D.Sc. (Tech.) Sakari Tervo

Aalto University
School of Electrical Engineering

AALTO-YLIOPISTO
SÄHKOTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Perttu Laukkanen

Työn nimi: Tilaimpulssivasteen ja auralisaation soveltaminen äänitarkkaamoakustiikassa

Päivämäärä: 24.2.2014 Kieli: Englanti Sivumäärä:8+88

Signaalinkäsittelyn ja akustiikan laitos

Professuuri: Akustiikka ja Äänenkäsittely Koodi: S-89

Valvoja: Prof. Tapio Lokki

Ohjaaja: TkT Sakari Tervo

Työn sisältö jakautuu kahteen osakokonaisuuteen. Työn ensimmäisessä osassa sovelletaan tila-aikavastetta äänitarkkaamoiden varhaisen äänikentän tutkimiseen. Työssä mitattiin 13 huoneen tilaimpulssivasteet, sekä paikannettiin varhaiset akustiset heijastukset kuvalähdehajotelma-menetelmällä. Tila-aikavasteiden avulla tarkasteltiin äänitarkkaamoiden varhaista äänikenttää, ja huomattiin visualisointien kuvaavan intuitiivisesti tarkkaamoiden akustisia ominaisuuksia sekä tuovan tehokkaasti esille eroja eri tarkkaamoiden välillä. Heijastusten tarkempaa tutkimista varten toteutettiin käyttöliittymä, jolla pystyttiin tarkemmin visualisoimaan varhaisia heijastuksia eri ajanhetkillä.
Toisessa osassa työtä jokaisen mitatun tarkkaamon äänikenttä koostettiin kuvalähdehajotelman perusteella kaiuttomassa huoneessa, ja tarkkaamoita vertailtiin kuuntelukokein. Kuuntelukokeisiin osallistui 15 äänituotannon ammattilaista, joista 12 oli miksaajia ja 3 masteroijia. Koehenkilöiden tehtävänä oli ensin parivertailun avulla selvittää mikä tarkkaamo parhaiten sopisi heidän työhönsä ja toiseksi kertoa perustelut valinnoilleen haastattelussa. Lopuksi tarkkaamoita kuunneltiin yksitellen, ja koehenkilöt kuvailivat omin sanoin jokaista tarkkaamoa.
Tulokset osoittivat, että keskimäärin miksaajat pitävät akustisesti kuivista tarkkaamoista, ja heille tärkeitä asioita ovat jälkikaiunta-aika sekä stereokuvan tarkkuus. Toisaalta masteroijat pitävät hieman kaiuntaisemmista tarkkaamoista, ja heille taajuusbalanssi on tärkeä tekijä. Miksaajilla preferenssi näytti myös muuttuvan yleisesti siten, että kaiuntaisella musiikilla pidettiin hieman kaiuntaisemmista tarkkaamoista, ja kuivilla äänitteillä kuivista tarkkaamoista. Preferenssissä oli myös vaihtelua koehenkilöiden välillä.

Avainsanat: Auralisaatio, huoneakustiikka, studiotarkkaamo, tilaimpulssivaste, kuvalähdehajotelma

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Perttu Laukkanen

Title: Evaluation of Studio Control Room Acoustics with Spatial Impulse Responses and Auralization

Date: 24.2.2014 Language: English Number of pages:8+88

Department of Signal Processing and Acoustics

Professorship: Acoustics and audio signal processing Code: S-89

Supervisor: Prof. Tapio Lokki

Advisor: D.Sc. (Tech.) Sakari Tervo

The scope of this thesis is divided to two main parts. Firstly, this thesis presents the use of spatiotemporal visualizations to study the early acoustic reflections of studio control rooms. Spatial impulse responses of 13 rooms were measured and acoustic reflections tracked with Spatial Decomposition Method (SDM). Spatiotemporal visualizations of each control room-loudspeaker combination were then analyzed. In addition, a tool for more exact tracking of acoustic reflections was implemented. Spatiotemporal visualizations were found to intuitively describe the early sound field in control rooms, and connections between room treatments, geometry and spatiotemporal visualizations could be drawn.
Secondly, this thesis studies the preference of studio control rooms among professional mixing engineers. To enable A/B comparison between different control room-loudspeaker combinations, rooms were auralized on the basis of SDM-analysis. In the listening tests, auralizations were played back with 30-channel loudspeaker system in an anechoic chamber. The preference test was conducted in a form of pair comparison, where subjects listened to the auralizations, and chose the room in which they would prefer to work. After the preference tests, subjects were interviewed to reveal their arguments behind the preference decisions. Also each control room was listened to individually, and mixing engineers were asked to describe the acoustics of each room with their own words.
Finally, the results of the listening tests and their connections to spatiotemporal visualizations are discussed. The results of the listening tests clearly showed that mixing engineers prefer acoustically dry rooms. Accurate stereo image and the amount of room reverberation were the most important factors for them. In contrast, mastering engineers seemed to prefer more lively rooms and the frequency balance was the most important factor for them. The preference rating seemed also to vary between different music samples.

Keywords: Auralization, Small room acoustics, Studio control room, Room impulse response, Spatial decomposition method

Acknowledgements

The work for this thesis was done within the Virtual Acoustics research team at the Department of Media Technology in Aalto University School of Science. First of all I would like to thank Professor Tapio Lokki for the opportunity to research the topic and especially for the supportive attitude towards my work. I would like to offer my gratitude also to my instructor D.Sc. (Tech.) Sakari Tervo, who has taken significant responsibility for improving this thesis and making everything according to the necessary scientific accuracy. I would also thank Dr. Tervo for the significant contribution to the measurements and listening tests, but first and foremost for keeping the spirit up in the office and being an encouraging instructor.

In addition, I would like to thank Jukka Saarelma and Dr. Jukka Pätynen for the daily discussions and assistance with Matlab, Antti Kuusinen for comments related to the listening test procedure, and Ilkka Huhtakallio and Archontis Politis for their help with listening test practicalities. Also thanks to Professor Ville Pulkki for possibility to use Aalto University Department of Signal Processing and Acoustics' Multi channel chamber for conducting the listening tests. Finally I would like to direct my thanks to every Finnish sound engineer involved in this work. Without your interest and contribution this study would not be possible in the scale it was realized.

Otaniemi, 24.2.2014

Perttu Laukkanen

Contents

Symbols and abbreviations

Symbols

c	Speed of sound
τ	Time-difference
f_s	Sampling frequency
f_c	Schroeder frequency
T_{60}	Reverberation time

Abbreviations

CID	Controlled Image Design
CSD	Cumulative Spectral Decay
DS	Direct Sound
EDT	Early Decay Time
ETC	Energy-time Curve
GUI	Graphical User Interface
IR	Impulse Response
LEDE	Live-end, Dead-end
LTI	Linear and Time-invariant
RFZ	Reflection Free Zone
SDM	Spatial Decomposition Method
SPL	Sound Pressure Level
TDOA	Time Difference of Arrival
VBAP	Vector Base Amplitude Panning

1 Introduction

Studio control room acoustics plays a key role in the music production industry. Control room and loudspeaker setup are the final indicators for a mixing or mastering engineer to make decisions before a piece of music reaches the general public. If the control room is not properly acoustically treated, engineers cannot be sure whether they are finalizing a masterpiece or a messy mix with a poor balance. The control room treatment is also important in the recording stage, when the audio engineer decides, for example, which microphones to use and in which setup. For these reasons it is important to investigate the control room acoustics to better understand the causes behind different acoustical phenomena and especially their perceptual relevance.

However, most earlier research on studio control room acoustics, and especially studies related to room reflections, are limited to measurement and analysis of monaural impulse responses [28, 38, 61]. Monaural impulse response can illustrate that reflections are coming at certain time instances without solving their directions. In addition, monaural impulse response is used to solve the frequency response of a loudspeaker–room combination in a certain receiver position.

Several intensity-based methods in order to visualize three dimensional sound fields in small rooms (rooms with surface area from 10 m^2 to 100 m^2) are presented in [23, 43]. These preliminary studies used laborious intensity measurements to acquire 3D-visualization of reflected sound energy. Sound energy was visualized using colored bullets and different opacity of the bullet denoted the amplitude of the reflection. No magnitude scale was included in visualizations, thus they lacked intuitiveness and detail.

Also the subjective attributes of control rooms have been formerly studied only based on listening tests made individually in different control rooms, or only trying to reveal the effects of some particular acoustic properties at a time [26, 32, 56]. As far as the present author knows, any A/B comparisons between different control rooms have not been conducted.

This thesis proposes to study the early acoustic reflections in the control rooms by visualizing them spatiotemporally as well as auralizing sound fields to enable the A/B comparison of room-loudspeaker combinations between different control rooms.

The primary aim of this thesis is to solve the acoustic properties of an optimal control room on the basis of both objective spatiotemporal measurements and listening tests. Also the attributes that are the most important for sound engineers are tried to specify. The secondary purpose of this work is to examine the accuracy of the utilized methods, formerly used in concert hall studies, and find out if they reveal reliable data also from small rooms. As a by-product, this thesis also reviews the acoustic conditions in eight control rooms where a major part of the music is mixed and mastered in Finland. Also the possible acoustics defects detected with measurements are revealed, and physical reasons behind them discussed. "There is something strange in this room, but I cannot specify what it is." is not an uncommon phrase from mixing engineers. This kind of myths are tried to solve in this thesis.

The thesis is organized as follows. Chapter 2 presents key concepts related to recording studios and room acoustics in general. Chapter 3 covers previous research in control room acoustics, including different control room design principles, research based on objective measurements and research based on perceptual and subjective studies. Chapter 4 presents the measurement and analysis methods of this study. Chapter 5 reveals the obtained results and Chapter 6 discusses about the results and proposes possible future work. Finally, Chapter 7 concludes the thesis.

2 Background

This chapter presents background information needed to understand the rest of the thesis. Chapter 2.1 presents different parts of music production facilities. Chapter 2.2 introduces key concepts related to room acoustics, and finally Chapter 2.3 gives a walkthrough about a typical record production process.

2.1 Music production facilities

2.1.1 Recording studio

Recording studio is a facility for sound recording and mixing. Typical recording studio consists of a room called the studio or live room, where instrumentalists and vocalists perform, and the control room, where sound engineers operate. In control room, there is often also a producer piloting the recording process with the sound engineer. In addition to a large live room, there are often smaller rooms called isolation booths to accommodate loud instruments such as drums or electric guitar amplifiers. There is usually a machine room for a computer and a small booth for vocals in the immediate vicinity of the control room. It is practical to have a line of sight from the control room to studio as well as to the vocal booth. Windows are the common solution to enable the line of sight, but in these days screen and camera combinations are also used. Figure 1 illustrates typical floor plan of a commercial recording studio. [3]

2.1.2 Control or Mix room

Control room or Mix room is the space where the recording engineer monitors the sound during the recording process. The same room can be used also as a mixing room, where multiple tracks are assembled into a final 2-channel or surround sound product, i.e., a mix. In a large music recording facility, mixing may be performed in a separate, dedicated room, but acoustical properties of control and mixing rooms are typically close to each other. [3]

2.1.3 Mastering room

Mastering room is the facility for final check for the measures, e.g, balance and dynamics of the music piece. Mastering engineer considers these issues, and may take necessary actions to change them, but he may also recommend re-mixing certain portions before recording is released. Mastering engineer also takes the final responsibility to take full advantage of the medium, i.e., vinyl, CD, or digital audio file, in which the piece will be released before it is sent to production facility. The acoustic properties of mastering rooms are usually quite similar to control or mix rooms, but some mastering engineers prefer more reverberant rooms than what the mixing rooms are on average. It is also common that mastering engineers have specific personal preferences related to their room or loudspeakers. [3]

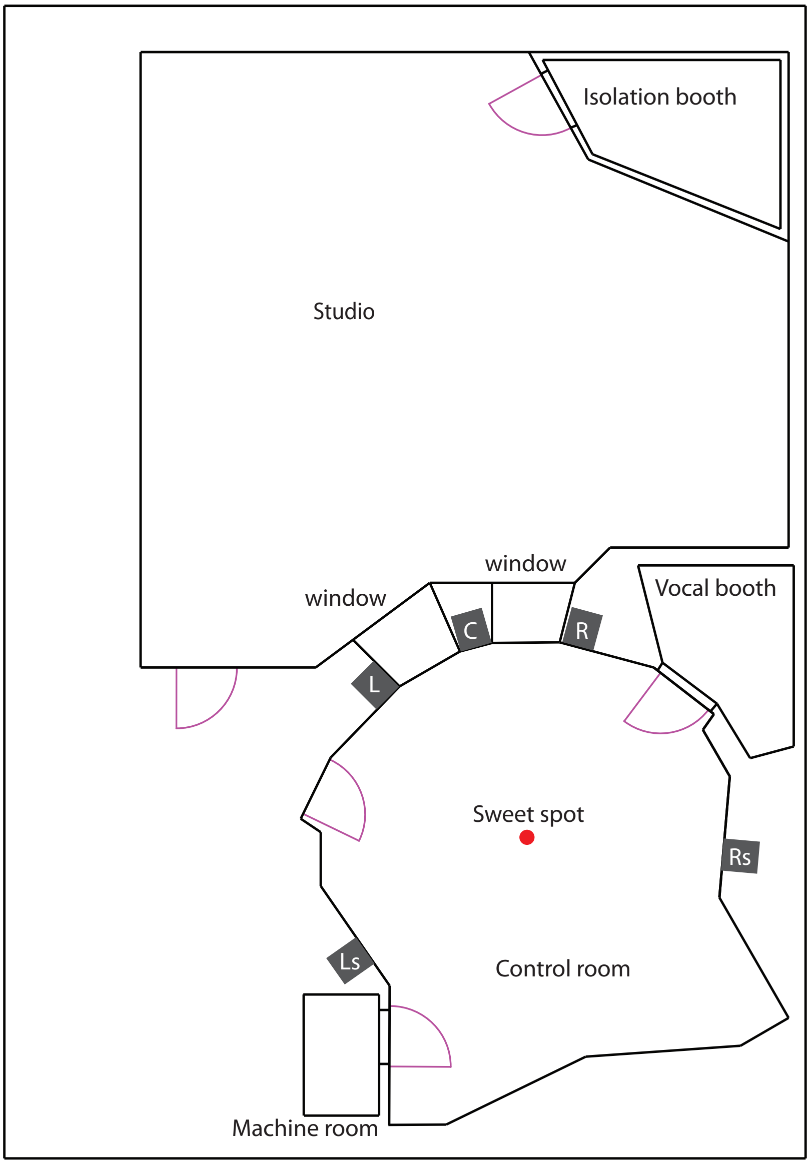


Figure 1: Floor plan of a typical recording studio complex. L, R, C, Ls and Rs presents Left, Right, Center, Left surround and Right surround loudspeakers according to ITU-R recommendation [27].

2.1.4 Production room

Commercial recording facilities often include small production rooms which may be rented to independent producers or music composers. Even more common for the composers is to set up work spaces in their homes. These rooms are often quite small and designed to give the artist efficient work flow. Today, even a lot of final mixing is done in these production rooms or home studios. Acoustic treatment in these rooms may be limited and almost invariably close field monitoring is used instead of flush mounted large loudspeakers. [3]

2.1.5 Sweet spot

The sweet spot or the listening spot is a term to describe the optimal area between two speakers where sound engineer is fully capable of hearing the stereo audio mix. In the case of surround sound, this is the focal point between four or more speakers. In this work, the receiver position R1 is in the sweet spot in every measured room. The red dot in Fig. 1 illustrates the position of the sweet spot in surround control room.

2.2 Room acoustics

Sound in every space is a mixed combination of a direct sound from sound source and the reflected sound from the boundaries of the room. The perception of the reproduced sound in a room depends decisively on the division between the direct and the reflected sounds. This chapter presents key concepts needed to understand sound propagation in small rooms.

2.2.1 Absorption

Absorption converts sound energy into heat. When a sound wave collides with some typical boundary in a room, say a fitted carpet, part of its energy is absorbed by the surface and the rest is reflected. How much of the sound energy is absorbed depends on the absorption characteristics of the surface. Absorption characteristics are described by absorption coefficient α which can have value between 0 and 1, so that material with $\alpha = 0$ reflects all sound energy, and material with $\alpha = 1$ absorbs all sound energy. Typical absorptive materials used in small rooms are porous panels and their absorption coefficients increase as the frequency increases. Total absorption area of a room consisting of different materials can be calculated as follows:

$$A = \sum_{i=1}^{n} \alpha_i S_i = \alpha_1 S_1 + \alpha_2 S_2 + ... + \alpha_n S_n, \tag{1}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the absorption coefficients of each surface material in the room while $S_1, S_2, \ldots, S_n$ are the surface areas corresponding the surfaces having the same subscript value as those for absorption coefficients [51].

2.2.2 Specular reflections

Specular reflections are mirror-like reflections, where the sound incidence and the reflection have the same angle with respect to the surface normal. If the surface curves are much larger in scale than the wavelength of the sound this rule holds locally at each point along the surface. Thus, specular reflections follow Snell's law [65] and can be modeled as rays or image-sources [2, 25]. In many control room design principles, early specular reflections occurring in the listening position are advised to absorb, scatter or reflect, as they are found to introduce audible colorations in perceived sound [17, 56, 60].

2.2.3 Scattered reflections and diffusion

If the surface faced by sound is rough, on the scale of, or somewhat larger, than the wavelength, the reflection becomes diffuse. Thus, the sound wave is reflected to various directions, not just in the specular direction. Figure 2 illustrates the difference between a specular and a diffuse reflection.

The sound field in a volume is defined as a perfectly diffuse when directional energy density is equal for each point and direction. In other words, the average sound pressure level is equal in each point within the space, and sum intensity is zero over an arbitrary plane in the space. Sound field can be nearly diffuse in a large concert hall, but in a small room diffuse sound field rarely exists [25, 51].

The ability of a diffusive structure to scatter different frequencies depends on its dimensions and conformation. Schroeder has studied extensively the mathematical sequences for designing the most effective diffusers [48]. In addition, a comprehensive review of the topic is written by Cox and D'Antonio [13].

2.2.4 Diffraction

Diffraction occurs when a sound wave encounters an edge. A practical example of this is a wave bending around a corner. The wave does not form a hard shadow and a person is able to hear sound around the corner. The amount of bending that occurs depends on the relation between the wavelength and the size of the wedge with which it interacts. The diffracted energy is proportional to the wavelength in a way that sounds with longer wavelengths diffract more than sounds with smaller wavelengths. [51]

2.2.5 Room modes

Room modes are a collection of resonances that exist almost in all practical enclosed spaces excited by an acoustic source such as a loudspeaker. In physics, resonance is the tendency of a system to oscillate with greater amplitude at some frequencies than at others. Room resonances occur at frequencies being related to one or more of the room dimensions. Axial mode frequencies for the rectangular room can be

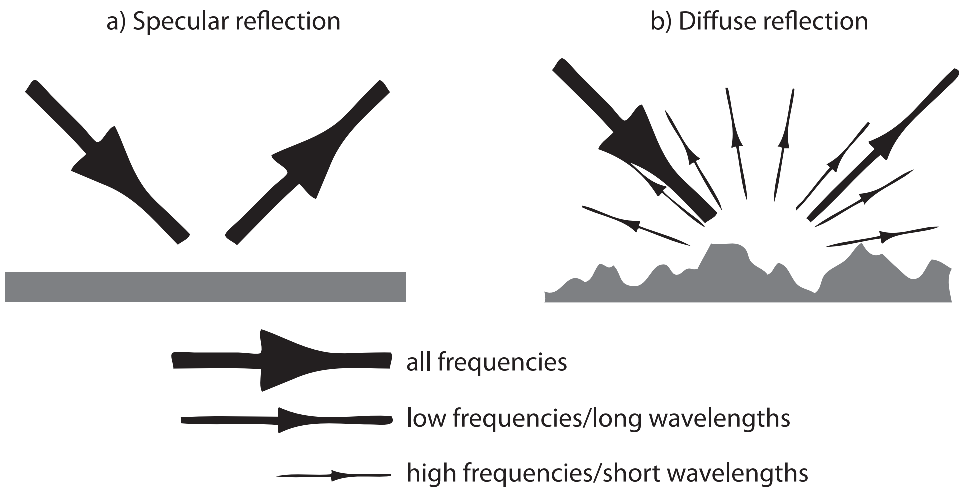


Figure 2: The basic concept of reflecting and diffusing surfaces. If the surface is completely rigid and flat, all frequencies are reflected specularly (a). If the surface is somewhat uneven (b), frequencies where the wavelength is long compared with the protruding features of the surface, there is still no scattering, and the sound is reflected as if the surface was flat. However, at frequencies where wavelengths are shorter in comparison with the surface conformation, sound is diffused in many directions simultaneously. Usually in small rooms, wideband diffusive structure can be too large and expensive to build, thus almost all practical diffusive structures are only partly diffusive like the surface in (b). Drawn after Toole [56].

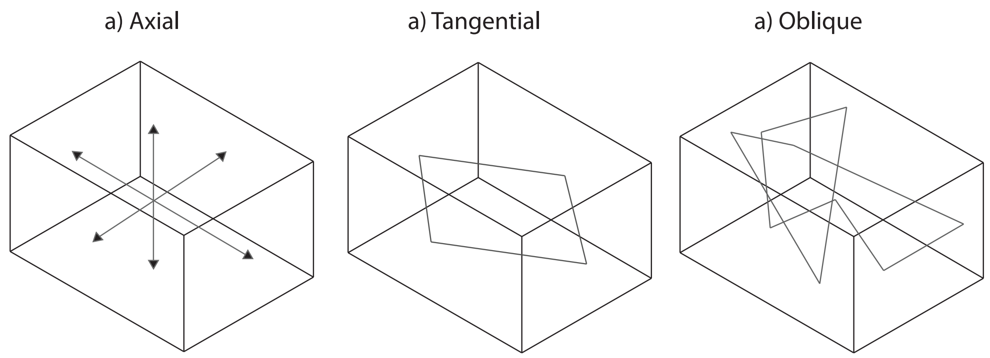


Figure 3: Three classes of room modes: (a) axial: length, width and height, involving two parallel surfaces, (b) tangential, involving four of the surfaces. (c) One of the many possibilities for oblique modes that involve all surfaces. Drawn after Toole [56].

calculated by:

$$f_{n_x n_y n_z} = \frac{c}{2}\sqrt{\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2}, \tag{2}$$

where $f_{n_x n_y n_z}$ is the frequency of the mode defined by the integers applied to dimensions x, y and z, l is dimension of the room in (m) and c is the speed of sound. For example $f_{1,0,0}$ is the first-order length mode (x dimension) and $f_{0,2,0}$ is the second-order width mode (y-dimension). In addition to axial modes, also tangential and oblique modes exist. Figure 3 illustrates three classes of room modes.

Input of acoustic energy to the room at the modal frequencies and their multiples causes standing waves. These standing waves results the level of the particular resonant frequency being different at different locations in the room, making the low-frequency distribution uneven within the room. Standing waves are usually the biggest problem in small rooms for reproduced sound, because controlling them needs large and heavy structures. In practical rooms the most problematic modes are usually axial at a relatively low order [25].

2.2.6 Reverberation time (T_{60})

Reverberation time (T_{60}) refers to the time period in which sound attenuates to inaudibility. Technically it is defined as the time when sound has attenuated 60 decibels after the sound source has stopped generating sound. Since the calculation of T_{60} is based on an assumption of exponential decay, it is not necessary to have a 60 dB of decay, but a smaller portion of the available dynamic range may be evaluated and the result simply scaled to correspond 60 dB of decay. T_{60} was first presented in 1898 by Sabine as an objective measure for studying room acoustics [7]. In control rooms, typical T_{60} values averaged within whole audio frequencies are 0.15 - 0.4 seconds, and in domestic listening rooms 0.4 - 0.6 seconds. An approximation of the T_{60} in an enclosed space can be calculated if the absorption area of the space is known:

$$T_{60} \approx 0.1611\frac{V}{A}, \tag{3}$$

where V is the room volume and A is the absorption area of the room. It is important to note that the standardized T_{60} measurement [1] assumes perfectly diffuse sound field, which does not exist in small rooms, as the sound field is usually dominated by standing waves and specular reflections. For this reason, the results of T_{60} measurements should be interpreted with care in the case of small rooms, if the exact documentation of the calculation is not presented [33]. In control room acoustics, T_{60} is often measured with own monitor loudspeakers of the control room, and this measurement does not correspond to one standardized in [1].

2.2.7 Early decay time (EDT)

Early decay time (EDT) is defined as the time required for the sound energy level to decay 10 dB after an excitation has stopped. To enable direct comparison to T_{60}, the result is scaled by a factor of 6. Compared with T_{60}, EDT is a measure concerning the early part of the decay process, thus it better describes the early reflected sounds in rooms. In a perfectly diffuse sound field EDT equals T_{60} [33].

2.2.8 Schroeder frequency

The Schroeder frequency, the critical frequency and the cut-off frequency are all synonyms and denote the approximate boundary between reverberant room behavior and discrete room modes [50]. In other words, below the cut-off frequency, the acoustic behavior of a room is dominated by separate room modes (modal region) and above the cut-off frequency, by a dense modal overlap with statistical properties (Schroeder region). Schroeder proposed a formula for approximating the critical frequency in a room:

$$f_c \approx 2000(T_{60}/V)^{0.5}, \tag{4}$$

where T_{60} is a reverberation time in seconds and V is a room volume in cubic meters. However, in small rooms Schroeder frequency is a mismatched concept, as it assumes meaningful reverberation times and a strongly diffuse sound field. That is why calculated f_c values are almost always too low in small rooms [5]. As the transition between modal region and Schroeder region is not sharp, the so called transition region [56] is often used to describe the transition area from purely modal region to Schroeder region. In small rooms, Schroeder frequency is typically in the region of 200 - 300 Hz.

2.2.9 Room impulse response

When a sound wave propagates in an enclosed space, it is affected by the phenomena mentioned above. The signal received in a sensor is therefore a modified version of the signal emitted by the source. If the source signal is a single Dirac impulse $\delta(t)$ [51], signal arriving to a sensor is called the impulse response (IR). If the system is Linear and Time-invariant (LTI) it can completely be characterized by its impulse response [36]. By knowing the impulse response of a LTI system, it is possible to calculate the output for an arbitrary input signal using a convolution integral.

$$y(t) = x(t) * h(t) \tag{5}$$

$$= \int_0^\infty x(\tau)h(t-\tau)d\tau, \tag{6}$$

where $y(t)$ is output, $x(t)$ input, and $h(t)$ impulse response of the system. Generating a Dirac impulse with a loudspeaker is unpractical, so room impulse response is usually measured, for example using pseudo-random noise [10] or logarithmic sweep [18] as a stimulus signal. Comparison of the different impulse response measurement techniques is presented in [53].

2.2.10 Initial time-delay gap (ITDG)

Initial time-delay gap (ITDG) is the time interval between direct sound (DS) and the first reflection [8, p. 27]. It was first defined by Beranek during his research in concert hall acoustics, but it is also used in small room acoustics and especially in research related to studio control rooms [16].

2.3 The record production process

To better understand the goals and practices related to control room acoustics, a brief walkthrough about the record production process is given. Following illustrates the common procedure of album recording in a modern computer aided studio. However, some recording engineers still use full analog signal chain, and recording and mixing is basically done at the same time, but this kind of approach is rare in these days.

First, in recording sessions, the sound engineer builds a recording setup to the live room and makes decision which microphones to use and places them to suitable arrangement related to instrumentalists or vocalists. After that, in a sound check, the engineer listens to the sound of the instruments and vocalists both individually, and also all at the same time. In this step, it is important to have a quick access from the control room to the live room for adjusting the microphone positions and re-positioning instrumentalists within the room to handle possible leakage sounds appearing to microphones. Also proper monitoring is important, because if mistakes are made in the recording stage, it can lead to a much additional work in the mixing stage or in the worst case, ruin the whole record. Recording sessions are often very intensive and stressful due to the limited time available, thus all inconvenience related to the monitoring accuracy should be minimized [29, 39].

After the primary recording sessions, some additional instruments or vocals are commonly recorded as overdubs. The whole recording process of a full-length album can take from one day to several months or even years, depending on the genre, artist and the professional requirements set for the final product. It should be noted, that it is common to record parts of an album in different studios, and by several recording engineers. Thus, some kind of consistency between different recording facilities is desirable [39].

After the recording process follows an editing stage. In editing, the sound engineer selects the best take or possibly combines multiple takes into a composite master take. If reasonable take does not exist, engineer corrects the player errors, such as timing mistakes or missed notes by hand. In editing, it is common to use automated timing and tuning tools, which are effective in general, but they may add minor clicks or timing errors to the processed track. These kind of errors needs very accurate monitoring to be noticed. The amount of editing depends greatly on a genre and the skills of performers, and commonly it can be said that the better the performers are, the quicker the editing stage is.

The next step is the mixing. As Izhaki defines in his book [29]: "Mixing is a process in which multitrack material – whether recorded, sampled or synthesized – is balanced, treated and combined into a multichannel format, most commonly two-channel stereo" [29, p. 4]. Izhaki also suggests a less technical definition: "A mix is a sonic presentation of emotions, creative ideas and performance" [29, p. 4]. There are many different approaches to a mixing, and practices depend on the genre, facilities and especially on the mixing engineer. However, the main task for a mixing engineer is to find the correct balance between separate tracks and to ensure that the mix sound as good as possible in all kinds of sound reproduction systems. To achieve that, engineer has to listen to the mix both at low and high sound levels. Control room should be neutral in frequency response, as well as in reverberation time. When building the spatial structure of the mix, the accuracy of the stereo image is very important, and the listening setup should be able to form a clear image both for width and the depth dimensions in stereo.

In mixing, individual instruments are equalized to sound neutral and free from disturbing resonances (ringing) and hums. To achieve a clear mix, an equalization should take into account the masking effects between instruments at the same frequency region. Thus, there must be space for each instrument in frequency spectrum. To accent certain instruments, also compression is common tool for a mixing engineer. Compression can be upward compression which makes soft levels louder or downward compression which makes loud levels quieter [29, p. 263]. In his book [29] Izhaki presents common questions mixing engineer can ask himself [29, p. 24]:

- How loud instruments are in relation to one another?
- How instruments are panned?
- How do the different instruments laid-out on the frequency spectrum?
- How far are instruments with relation to each other?
- How much compression was applied on various instruments?
- How long are the reverbs?
- How defined instruments are?
- How different mix aspects change as the song advance?

Last eventuality to enhance the sound or repair problems before the record is pressed is the mastering. Mastering engineer's experienced ears and a room are the final "audio microscope", as Katz phrases in his book [31]. If the mix is perfect, nothing is done in the mastering, but it is common to add minor equalization and compression to make the final master as playable as possible on a wide variety of systems.

At the end, it is worth to mention that despite record labels see very little marketing potential in production figures, it is very common to a major record label to pay huge amount of money to get a specific engineer to mix or to master an album. This reveals the fact that a mix certainly plays a huge role in the success of an album, and the majority of people appreciate sonic quality more than they will ever care to imagine [29].

3 Previous research in control room acoustics

This chapter reviews the previous work related to control room acoustics research. Chapter 3.1 presents different design principles. Chapter 3.2 presents objective measures previously used in control rooms studies. Chapter 3.3 discusses about the use of an electronic room compensation in control rooms. Chapter 3.4 covers the earlier perceptual studies, and Chapter 3.5 presents the common specifications for an ideal control room. Finally, Chapter 3.6 summarizes the previous research in control room acoustics.

3.1 Different design principles

In the following, five control room design principles from the early 40's to present are presented in approximately chronological order. The main goals between different control room philosophies are examined and methods how to fulfill these goals are presented. Previously, several collective papers, e.g. [3, 58], have been published where different control room designs are examined. However, it must be kept in mind that lots of control rooms during the history were built without specific design theory by combining different principles. These approaches did not usually lead to very good end result, as they did not follow any specific principle [39]. As this work focuses on music production control rooms, multi-channel and cinema mixing preferences are left to minor attention.

3.1.1 Early designs from 40's to 60's

In the early days of recording industry in the 40's and 50's, control rooms were commonly very small and the only thing considered was to make rooms practical to work in. In the early 40's recordings were done directly to the disk and a single monitor loudspeaker was used to monitor the final recording. After the introduction of tape recorders in the late 40's, it was common to use three loudspeakers to monitor each channel in the tape recorder separately. In the 40's and early 50's there was no attempt to make control rooms to work as critical listening environments [3, 59]. In the late 50's Putnam designed facilities for United Recording Corporation in California [46]. These studios were probably the first to properly use stereophonic monitoring and where control room acoustics were considered. In 1957, the Capitol Tower was completed, but without stereo monitoring before the renovation of control rooms in 1959 [6, 47].

In the middle of the 60's when the number of channels increased in tape recorders, the summing of the recorded tracks for monitoring started to become common. In that time, also the idea that the control room should function as a reference listening room started to become widespread. [3, 39]

In 1966, British Broadcasting Corporation (BBC) standardized acoustical requirements for their broadcast control rooms. The goal was to make rooms not very dissimilar from the average conditions in domestic listening rooms. Thus, BBC control rooms that time were designed to have reverberation times of 0.4 seconds up to 250 Hz, decreasing to 0.3 seconds at 8 kHz. [3]

At the end of the 60's, thinking started to turn towards more revealing control rooms. In 1969, when Hidley designed several control rooms in the US, instead of trying to simulate living rooms, the goal was to provide more accurate monitoring. Hidley's goal was to provide an accurate stereo image for the entire working area at the mixing console. [3]

3.1.2 The Live-End, Dead-End (LEDE)

In 1979, Davis and Davis introduced a new approach to control room design called Live-End, Dead-End (LEDE) [17]. The main goal for the philosophy was to create a neutral monitoring environment, without making the room anechoic. Here 'neutral' means that the room does not add any colorations to the program material. Also the control room Initial Time Delay Gap (ITDG) was specified to be 2 - 5 ms longer than in the live room, enabling to observe the reflections of live room more accurately. These aims were achieved by making the front wall and ceiling of the room as absorptive as possible and the rear wall and rear side walls as diffusive as possible. Due to the absorptive front wall, early reflections from the loudspeakers were attenuated. As the rear and rear side walls are maximally diffusive, no specular reflections are arriving at the listening position.

LEDE concept was developed only for stereo listening and it cannot be applied to surround monitoring as such. Although the psychoacoustic theories behind the LEDE design have been much debated, especially the point behind the ITDG difference between control room and live room [58], the LEDE concept was widely used in 1980's and lots of commercially successful pieces were mixed in LEDE rooms in the 80's and 90's.

3.1.3 Reflection free zone (RFZ)

Reflection free zone (RFZ) principle is more like an elaboration for implementing the LEDE control room. It was introduced by Peter D'Antonio and John Konnert from RPG Diffuser Systems Inc. in 1984 [14]. The dead-end is achieved by flush mounting the monitor loudspeakers as close to a trihedral corner as is physically possible to minimize comb filtering effects in the low frequency response. Side walls and ceiling are made as absorptive as possible, as Davis proposed in [17]. The rear wall diffusion is realized by using quadratic residue diffusers.

3.1.4 The Non-Environment (N-E)

The non-environment (N-E) control room was first introduced by Hidley and Newell in 1991 and a review of it can be found in [39]. The idea behind the N-E room is to create yet more neutral room than the LEDE. This is achieved by making the ceiling, rear wall and side walls highly absorptive and leaving front wall reflective. Monitor loudspeakers have to be flush-mounted to the front wall, so the sounds coming from the loudspeakers are not reflected because loudspeakers are radiating away from the front wall. As the floor and front wall are hard and reflective, it makes natural ambience to the room and thus prevents the environment being too dry to work comfortably [39].

N-E rooms are generally slightly drier than LEDE rooms, which has given rise to the debate if the N-E rooms are too anechoic to work or if mixes translate worse to domestic listening rooms because of the significant T_{60} difference. However, Newell states that a common opinion among sound engineers is that control room should be noticeably drier than domestic listening rooms on average. Due to the absence of reverberation, room to room variations between different N-E control rooms are minor. This enables engineers to mix in different N-E rooms without thinking how mixes will translate to other room.

As LEDE rooms, N-E rooms are designed only for stereo listening and they do not work with surround monitor setup as such. Due to the need of large amount of wide band absorption, N-E rooms are quite expensive to build and that is why very few have actually been built [39, 58]. However, there have been many well-known mixing engineers that have preferred and supported N-E rooms, for example George Massenburg [39].

3.1.5 Controlled image design (CID)

In 1993, Walker from BBC introduced a new approach to the design of stereophonic control rooms called controlled image design (CID) principle [60]. Unlike earlier designs, CID theory did not apply huge amount of wide band absorption to prevent early reflections, but used non-absorbent surfaces to direct early reflections away from the listener. The design goal was to prevent energy components higher than -20dB relative to the direct sound to appear in the listening position during the first 20 ms from the arrival of direct sound. CID rooms were expected to be more pleasing to work due to the reverberation time closer to domestic living rooms.

When the prototype CID control room was built in BBC, Walker made measurements and noticed that the goal was not quite achieved. However, occupants considered the room to be live, but also quite accurate and revealing, which encouraged Walker to build three more CID rooms and design goal was adjusted to prevent energy components higher than -15 dB during the first 15 ms to appear. In 1994, Walker presented the measurement results of three other control rooms that were built in BBC's Broadcasting house, London, United Kingdom [61]. On the basis of measurements Walker concludes that the design goals were adequately achieved. Early reflections from direct sound to 15 ms after direct sound were attenuated at least 15 dB and rooms were accepted by sound engineers.

CID design was used reportedly mainly in BBC [58] although it contains two tempting viewpoints; firstly, not requiring huge amount of absorption, which takes lots of space and is expensive, and secondly, reverberation time being closer to domestic living rooms. The main drawback in CID design is that the area of the sweet spot is quite narrow, and thus design is most suitable for music production with few occupants [63].

Walker extended his CID principle to multi-channel control rooms in [63]. However, multi-channel listening required some simplifications to the design. As the

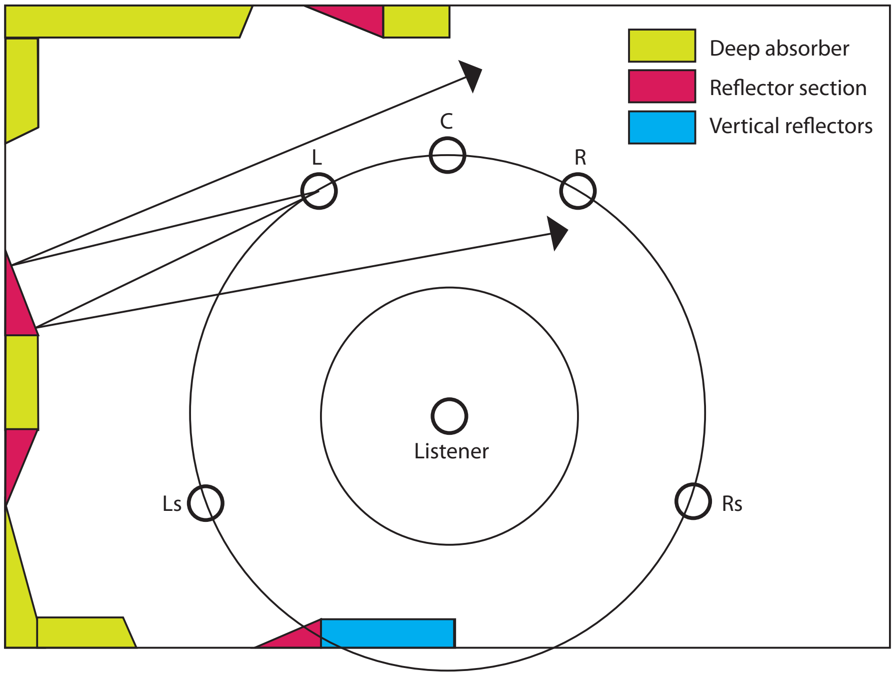


Figure 4: Layout of the wall structure for the left half of multi-channel CID room. Inner circle illustrates exclusion area free from early reflections. Drawn after Walker [63].

amount of loudspeakers increases, also the number of potential combinations of source and reflecting surface increases, which forces to make compromises when targeting reflective panels. Figure 4 illustrates completed wall design for half of the multi-channel CID room. In multi-channel CID rooms, Walker used diffusive structure in rear wall to redirect the reflections from all loudspeakers (blue color in Fig. 4).

3.1.6 MyRoom principle

In 2010, Petrovic and Davidovic proposed a novel desing approach for a control room, compatible for stereo as well as for surround monitoring [15]. The goal was to provide a better mix translation to other systems with less need for the engineer to adapt. As the expected scenario for the design was a small room, most probably located in a domestic environment, they named the design as "MyRoom" principle.

The MyRoom principle is very similar to George Massenburg's idea used in Blackbird Studios control room C which is based on wide band diffusion on all of the surfaces except the floor [9]. However, the specification of limited space caused that Petrovic and Davidovic could not directly apply the idea of Massenburg. The MyRoom principle used hybrid absorptive and diffusive treatment on walls and on ceiling. The idea was to absorb frequencies below 250 Hz and scatter frequencies over 1 kHz. This was achieved with custom structure where air can pass through the diffuser allowing low frequencies to absorb to the trap behind. Structure is

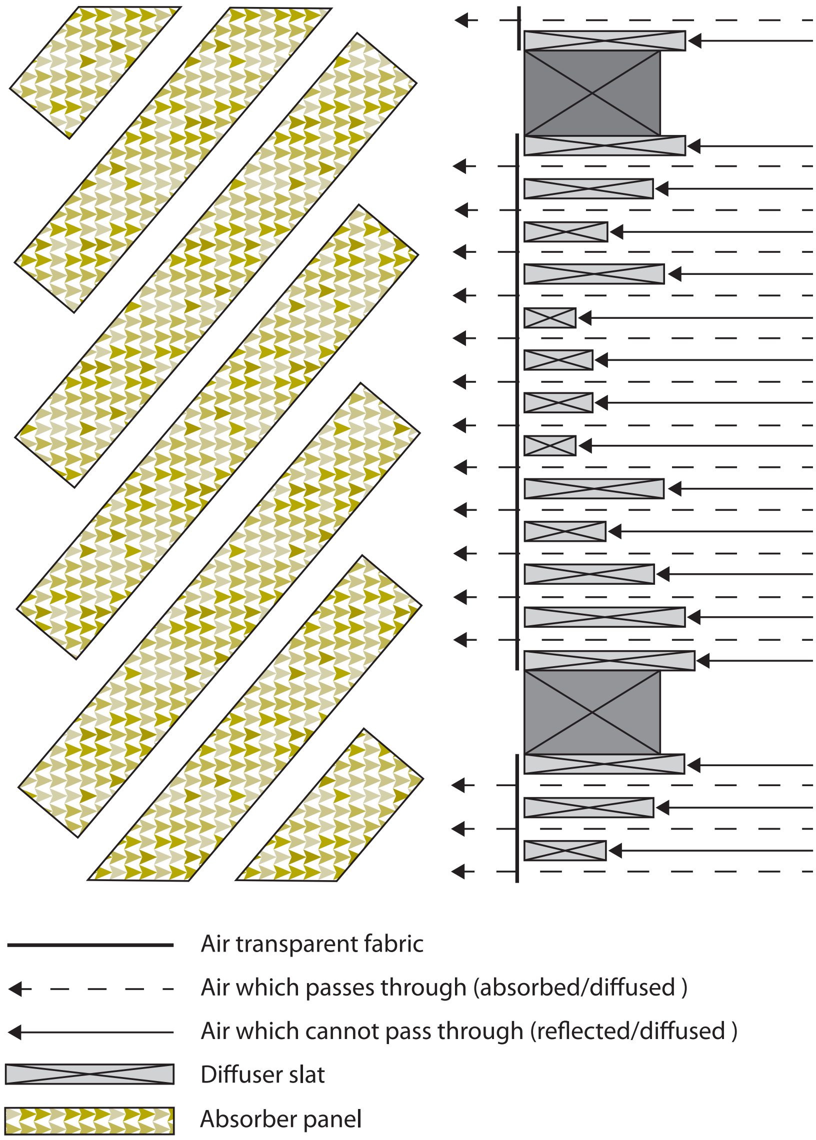


Figure 5: Hybrid diffusive and absorptive wall structure used in MyRoom principle. Drawn after Petrovic [15].

illustrated in Fig. 5. As the whole room except the floor is treated the same way, it gives the room premise to success in both stereo and surround.

According to the measurements and subjective response, two control rooms built under MyRoom principle, succeeded better than expected. Concluding results from [15], first reflections after direct sound were attenuated at least 20 dB in both rooms. Also the magnitude response was within limits of ± 4 dB from 40 Hz to 15 kHz, except in the 70 Hz area in a second room where the magnitude response was slightly more attenuated. Magnitude responses were measured all furniture in their places and without any electronic equalization before the loudspeakers. Reverberation times were as low as 0.15 - 0.20 seconds but rooms were still not perceived too dry. Instead, due to the large amount of diffusion, rooms were perceived to be acoustically bigger than their physical size. Detailed analysis of the design and results can be found in [15].

3.2 Research based on objective measurements

Control room acoustics and small room acoustics in general have been studied a lot and many objective measures are used to describe and rate the acoustic properties of rooms. Fundamental books where small room acoustics are reviewed are Newel's "Recording studio design" [39] and Toole's "Sound Reproduction, Loudspeakers and rooms" [56]. These authors have also published a large number of scientific articles about the topic. In the following, the most common objective measures used in small room acoustics research are presented.

3.2.1 Monaural impulse response

A monaural impulse response, and its visualization as an energy-time curve (ETC) have undoubtedly been the most used measure alongside the magnitude response in the study of control room acoustics, as well as in investigations related to loudspeakers. An impulse response reveals the complete acoustic characteristics of the measured system in a certain receiver position. The power of impulse response measurement is that other more revealing acoustic measures can be calculated from the impulse response data. The most common measures used in small room acoustics consist of the energy-time curve (ETC), frequency response and modulation transfer function (MTF).

From an impulse response graph, the amplitudes of individual reflections can be tracked in different time moments in a reasonable accuracy. Figure 6 (a) illustrates a normalized impulse response obtained from a control room. Measurement was taken in the sweet spot using the left stereo loudspeaker as a sound source. Some reflections can be seen within the first 3 ms and at approximately 8 ms. However, from a sheer impulse response graph, the cause behind the reflections can only be speculated.

It must be noted that in room impulse response measurements, the whole signal path, from the excitation signal to the measurement microphone and further to the measurement device (computer, audio analyzer), is included in the measurement. Thus, the measured impulse response includes the responses of loudspeaker, room, microphone, all cables and electronics in between. This must be taken into account and impact of all these parts have to be known when performing room impulse response measurements.

3.2.2 Energy-time curve (ETC)

Energy-time curve is a graph of sound energy with respect to time and it illustrates how sound decays within specified time. Energy in decibels can be calculated as

$$E(t) = 10\log_{10}(p^2(t)), \tag{7}$$

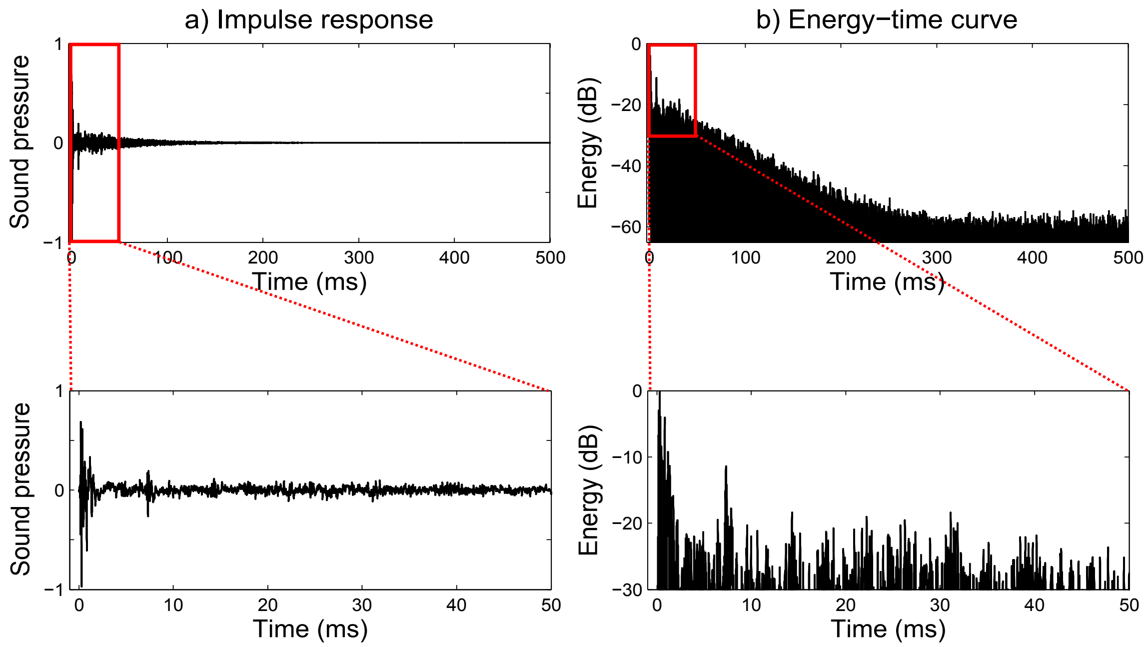


Figure 6: Example visualization of an impulse response (normalized sound pressure over time) (a), and energy-time curve (b) of control room, measured in the sweet spot using left stereo loudspeaker as the sound source.

where $p(t)$ denotes the sound pressure at time t. As seen in Fig. 6 (b), ETC visualizes individual reflections and their amplitude more clearly than the raw impulse response graph. From the shape of the ETC curve, specular and focused reflections can also be seen. If the decay in ETC is linear, smooth and lacks sharp peaks, the sound field is said to be free from specular and focused reflections. This is rarely realized in small rooms. For example in the ETC in Fig. 6 (b) there are many peaks in the decay. The ETC curve is used to a great extent in studies related to room acoustics and acoustic reflections, and especially measuring design goals, such as, "all reflections must be attenuated 15 dB or more during the first 15 ms" [15, 39, 56, 61].

3.2.3 Frequency response

A frequency response of a system can be obtained from an impulse response $h(t)$ via the Fourier transform:

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}dt. \tag{8}$$

The absolute value of the complex frequency response $H(f)$ is called the magnitude response, and the argument is called the phase response. As the name indicates, the frequency response reveals the output of the system at each frequency. In other words, the frequency response reveals the magnitude spectrum and the phase spectrum of a system. In this work, the system examined is a room and loudspeaker combination. In non-scientific literature and spoken language, the magnitude response is often mistakenly called as the frequency response.

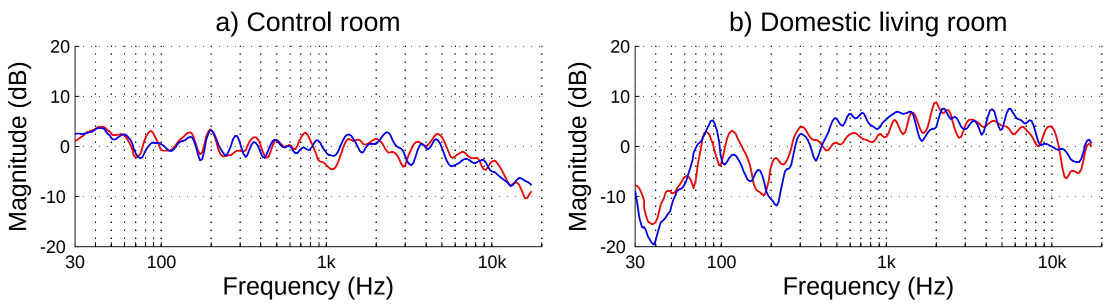


Figure 7: Example visualization of 1/3 octave smoothed magnitude response measured in control room (a), and in domestic living room (b). Red denotes left stereo channel and blue right channel.

Magnitude response plots are often smoothed to better correspond to the frequency separation accuracy of the human auditory system. Smoothing corresponds to averaging over certain frequency bands. A commonly used frequency bands in acoustics are the $1/n$-octave bands and 1/3-octave smoothing is found to best match to a human hearing [30, 56]. However, it can be useful in some cases to explore magnitude response in more detail, so, for example, 1/12-octave smoothing is also popular [55]. In this work, 1/3-octave smoothing is applied for all visualizations of the magnitude response.

Generally, a smooth, i.e., a flat magnitude response over the whole audio frequency range is the goal for an optimal room response [55]. However, this is rarely achieved in practice even in the most expensively built control rooms. Deviations from the flat magnitude response, i.e., colorations occur in almost every practical control room [56]. However, if the magnitude response fluctuation is within the limits of ± 4 dB in the whole audio range, frequency balance is considered uncolored. Figure 7 illustrates the magnitude responses of control room (a) and a domestic living room (b). In 7 (a) magnitude response is somewhat flat, thus it has no significant colorations. In contrast, in 7 (b) there is almost 15 dB dip around 200 Hz, which is probably due to room modes, and can impede proper perception of that frequency. In addition, Fig. 7 (b) shows significant magnitude difference between left and right channel at frequencies from 200 Hz to 2 kHz. This undoubtedly has effect to the stereo image. In contrast, magnitude responses in Fig. 7 a) are very similar, which tells about symmetry between left and right channels.

3.2.4 Cumulative spectral decay (CSD)

Cumulative spectral decay (CSD), commonly used interchangeably with 'Waterfall plot' is a common measure when studying time-frequency characteristics of a system. In room acoustics, CSD is used especially to study low frequency room modes. CSD is constructed by windowing the impulse response $h(t)$ backwards in time domain with a modified rectangular window, and applying the Fourier transform to the windowed response at each time step [12]. A CSD-plot is formed by displaying the magnitude response of $H(t)$ at the selected time steps. The x-axis commonly

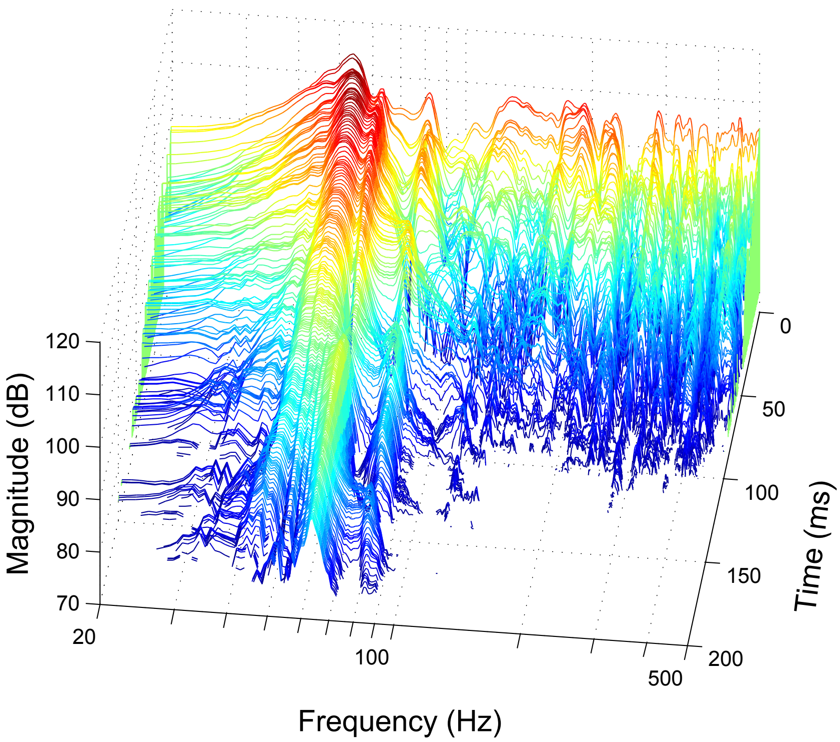


Figure 8: Example of the visualization of Cumulative spectral decay (CSD), i.e., 'Waterfall plot'. Note that for better frequency resolution, visualization includes only frequencies from 20 Hz to 500 Hz.

correspond frequency, y-axis magnitude and z-axis time.

It must kept in mind that the resolution of the CSD is much better at higher frequencies than at low frequencies. This is due to the constant bandwidth of the Fourier transform analysis. Figure 8 illustrates an example visualization of CSD. In Fig. 8 it can be seen significant resonant behavior at 50 - 60 Hz, which is most probably due to a room mode.

3.2.5 Modulation transfer function (MTF)

The modulation transfer function (MTF) measures the ability of a system to maintain the depth of modulation from an input signal at specific frequencies. MTF can be obtained directly from the impulse response [49]. The depth of modulation for each frequency can be calculated as follows:

$$m(F) = \frac{\int_0^\infty h_f^2(t) e^{-j2\pi F_t} dt}{\int_0^\infty h_f^2(t) dt}, \tag{9}$$

where h_f is the band-limited impulse response with center frequency f and F is the modulation frequency. For a discrete impulse response, Eq. (9) is given as:

$$m(F) = \frac{\sum_0^N (h_f^2(n) e^{-j2\pi F_n/f_s})}{\sum_0^N (h_f^2(n))}, \tag{10}$$

where N is the total length of the impulse response in samples and f_s is the sampling frequency. When calculating MTF values for separate frequency bands, care must be taken to ensure that digital filtering does not distort the scores. When using long digital filters, the filter response itself can dominate the MTF result, so that the measured MTF is mainly from the filter response rather than from the measured system [20].

Fazenda *et al.* have studied the use of MTF as a measure of the room–loudspeaker low-frequency performance [20, 22]. Their investigations include examination how MTF relates to room modes and further to the change in room volume, aspect ratio and damping. They also studied how MTF scores correlate to low frequency perception. Conclusions show that MTF does not correlate very much with the room volume. Results even show that smaller volumes give better MTF values, which is against the common belief that larger rooms will give better modal distribution and thus better low frequency accuracy. Neither the different aspect ratios do give much difference in MTF values. Instead, the room damping correlates highly with MTF values, low decay times giving the best MTF score. The overall conclusion is that MTF is useful measure of the accuracy how low frequencies are perceived and that the MTF plots are good indicators of which frequency range may be the most problematic in a listening room.

3.3 Electronic room compensation

Electronic room response compensation has dramatically increased in control rooms during the recent years. In addition to dip switches for bass level control, also more sophisticated methods for correcting the frequency response in the listening position have been studied and implemented [11, 37, 62]. Room compensation is mainly used at low frequencies to handle the frequency response anomalies caused by room modes.

Digital signal processing makes it possible to create adaptive filters, which can model very accurately the inverse responses for correcting frequency response anomalies as well as defects in a transient response. However, the fundamental problem in these corrective filters is that when they can give an almost absolute correction in the amplitude and the phase in one point in the room, they cannot make the absolute correction over a large area. All corrections at one point are gained at the cost of response anomalies elsewhere [39, 62, p. 350]. That is why compromises have to be done when equalizing room responses with these filters. However, if strong room modes exist, carefully made equalization can considerably improve the low frequency performance.

3.4 Perceptual studies

3.4.1 Toole *et al.*

Toole *et al.* have contributed significantly to the research of sound perception in small rooms [42, 55, 56]. They have striven to solve the relationship between objective

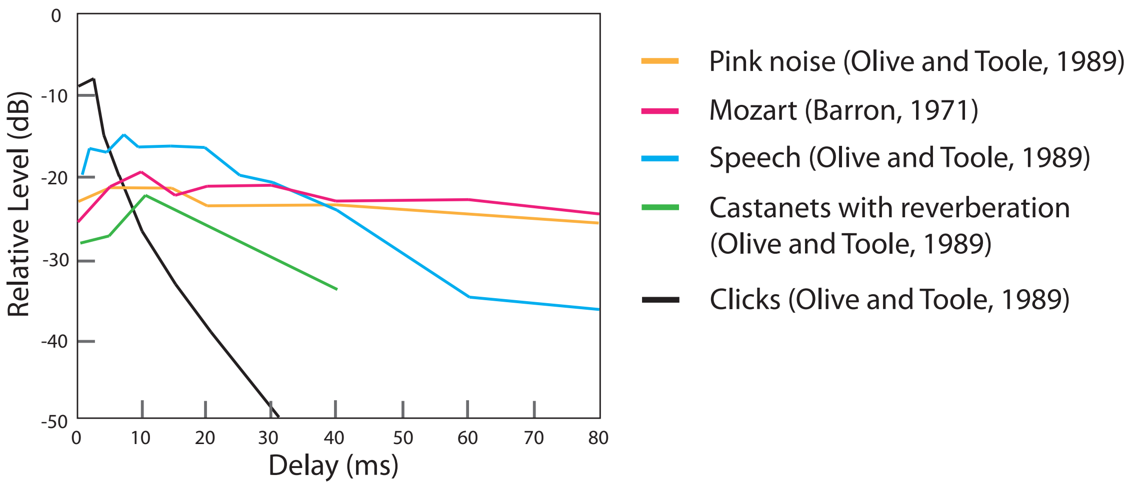


Figure 9: Detection thresholds for a single lateral reflection. Different colors represent results from different studies conducted with various test signals exhibiting different degrees of temporal extension (continuity). All studies are conducted in anechoic chamber. Drawn after Toole [56].

measurements and perceptual effects in small room acoustics. However, as Toole states in [56] and [55], a lot is yet to be done to completely understand sound perception in small rooms. Toole *et al.* have studied sound perception both from the entertaining point of view, as well as from the perspective of critical listeners. Thus, a large part of Toole's research is relevant to this thesis.

The most interesting results from Toole's experiments are the thresholds for different perceptual cues for reflections coming from different directions in different time delays. Large part of the listening tests have been done with speech, but some studies also with music. Figure 9 illustrates the results of different studies in the reflection perception. All studies were conducted in anechoic listening conditions, with slight variations in the horizontal angle for the lateral test reflection. From Fig. 9 it can be picked that with pink noise and classical music, the perception threshold for a single lateral reflection is approximately -20 dB related to the direct sound during the first 50 ms. With clicks, threshold is as high as -10 dB within first 3 ms, but drops dramatically after that.

Figure 10 illustrates results of Barron [4] drawn after Toole [56, p. 87], where the perceptual effects of single lateral reflection arriving from 40° to the side was studied. Listening tests were made with classical recordings (Mozart) [42]. Figure 10 illustrates that lateral reflections coming within the first 10 ms causes image shift towards the direction of the reflection. In control room, this means widening of the stereo image and thus enhancing the sweet spot size if the room is symmetric. The amount of widening required for the stereo image is a matter of opinion. Compromise has to be done between the stereo image accuracy and the widening of the stereo image, e.g., size of the sweet spot. From Fig. 10 it can be also seen that lateral reflections within 10 - 35 ms introduce tone coloration regardless of the level related to the direct sound.

Other notable point in Toole's experiments is that the sequence of several low-level reflections and a large single reflection were observed almost equally loud [56, p. 91]. The message here is that it could be misleading to assume that if large reflecting surfaces are broken on the basis of impulse response measurements, the audible effects of reflection will be reduced. This kind of effect discloses the persistent problem in a relation between measurements and psychoacoustics, as human perception is usually nonlinear while measurements are linear.

Considering room reflections, Toole concludes that the reflections from front and back do not have any positive effects on the listening preference. Instead, early lateral reflections from the side seemed to contribute in a positive way at least in entertaining purpose. It can be also concluded from the study of Imamura *et al.* [26], that absorption of the first reflections on the side walls causes "the width of sound image" to be narrower and "Envelopment" to be lower. Absorption of the first reflections on the front wall and ceiling make "the width of sound image" narrower and "Clarity" increase. Absorption of the first reflections on the back wall also makes "the width of sound image" narrower and "Clarity" increase.

In [56, p. 177], Toole presents results of a massive listening test conducted by Olive *et al.* [41], where three different loudspeakers were subjectively rated in four different rooms. First experiment was conducted in a "live" manner, where listeners evaluated all three loudspeakers in one room before moving to the next one. Reproduction was recorded binaurally for each loudspeaker-room combination, and the same test was conducted using headphones. Results showed that a loudspeaker was highly significant and room was not a significant factor and that the results of live and binaural tests were essentially the same. From this it can be deduced that listeners adapted to a room and were able to judge the pure loudspeakers quite accurate.

In a second test, using the same binaural recordings, room-loudspeaker combinations were judged in different context. Now listeners rated the same loudspeakers located in the same position between the four rooms. Thus, there were four comparisons per one trial. In the second test room became the highest significant variable whereas the effect of loudspeaker was not found significant, meaning that there were highly significant differences in preference due to the room factor. These results show that adaptation has a major effect when judging loudspeaker performance. However, this is not saying that there were no interactions between individual loudspeakers and individual rooms. There actually were, and especially at low-frequencies. It was also noticed, that in multiple comparison tasks, listeners tend to make judgements on a relative scale but they were less able to make consistent judgements on an absolute scale. Thus, test showed that the context in which comparison were made, influences listeners' preference ratings significantly.

3.4.2 King *et al.*

King, Leonard and Sikora have conducted several studies related to the effect of room reflections in critical listening [32] [34]. In [32], King *et al.* explored sound engineer performance when lateral first reflections were, minimized, maximized and

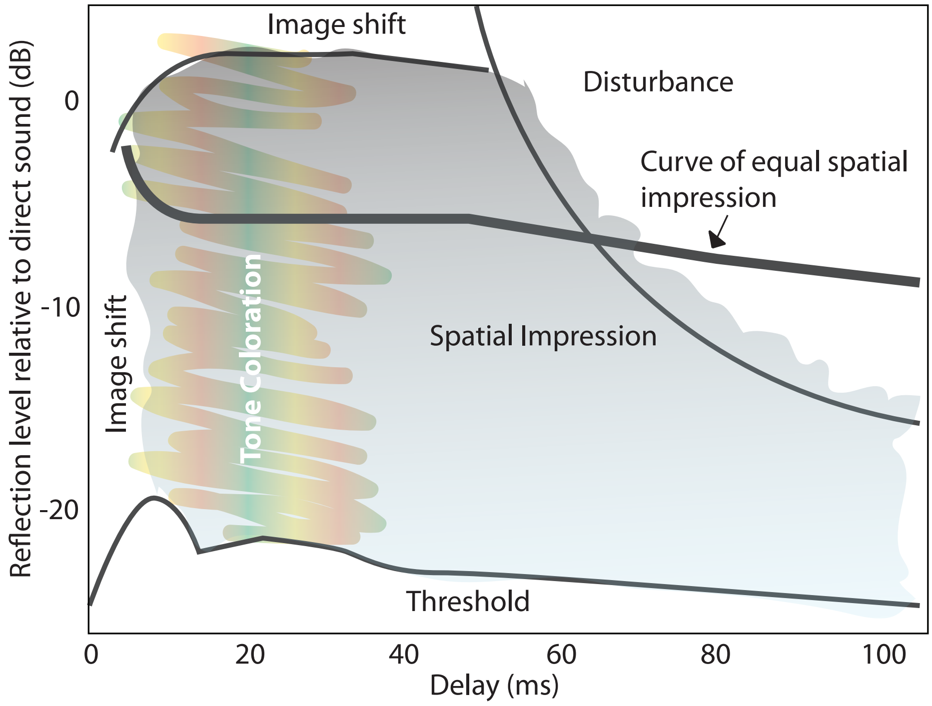


Figure 10: Subjective effects of a single reflection arriving from 40° side, including the effect of reflection level and delay compared with direct sound. Data is from Barron [4] and picture is drawn after Toole [56, p. 87]. Experiments were done with classical music. The lowest curve indicates the hearing threshold. Above this at short delays (less than 10 ms), listeners reported an image shift in the direction of reflection. At delays larger than 10 ms, listeners reported "spatial impression" where the source appeared to broaden and the music started to gain body and fullness. The spatial impression increased as the level of reflection increased, which is illustrated in the figure by the increased shading density. The curve of an equal spatial impression shows that for short delays, the reflections must be higher in level to produce the same effect. At high levels and long delays, disturbing echoes were heard, which is the upper right corner in the figure. At delays between 10 ms and 40 ms and at all levels, some tone coloration was heard (colored brush strokes in the picture). The areas identified as exhibiting an image shift refer to impressions that the principal image has been shifted toward the reflection image. At short delays, this would sum up the localization of reflection to the leading loudspeaker. At longer delays, the image would likely be perceived to be larger and less clear. Finally, at yet longer delays and higher sound levels, a second image at the location of the reflection will be perceived as an individual sound source. From the data of Barron, it is not clear where exactly these divisions occur.

diffused. Sound engineers were asked to adjust the level of a female soprano to the orchestral backing track in all these three conditions. The results shows that there were slight differences in the time needed for completing the task depending on the order in which different treatments were tested. However, after adaptation to the test procedure, normal performance was achieved. No significant difference was noticed in accuracy between different treatments. Conclusion was that due to the adaptation, no prominent difference in sound engineer performance occurred between three alternative side wall treatments. However, there is a limit to what we can adapt to, and as Toole speculates in [57], adaptation very likely utilizes some portion of our neural capacity and perhaps causes fatigue and stress. Thus, working longer periods of time in conditions that requires adaptation to a certain deficiency or a coloration could probably be unhealthy for mixing engineer. However, this kind of effects are still waiting to be proven and topic needs further research by neural scientists in collaboration with acousticians.

In [34], Leonard *et al.* studied the effect of a room in adjusting reverberation level in a mix. The task was to add a reverberation to a fixed stereo mix, first in a standard studio control room (averaged T_{60} of 0.2 s), and second in a highly reflective mix room (averaged T_{60} of 0.4 s). Results showed significant differences in reverberation levels set in each acoustical environment. The reverberation was mixed 1.32 dB lower in the reflective environment on average. Thus, the conclusion was that the room treatment had a significant effect to the reverberation level adjustment.

3.4.3 Fazenda and Davies

In 2001, Fazenda and Davies published a study where the aim was to identify the common language, views and preferences of professional sound engineers who commonly work in control rooms [21]. These opinions would give a studio designers an indication of where to best aim their efforts in order to improve control room designs. The survey was conducted in a form of semi-structured interviews, where 18 professionals were asked their opinions on control rooms they currently work or have worked in the past. Preferences and views regarding reverberation, stereo image, envelopment and the positioning of monitor loudspeakers were discussed. The survey was carried out in London and the North West of England.

Results were presented as quotes extracted from the interviews and the general trend identified for each parameter asked was indicated. Uneven low-frequency distribution within the room was experienced as the most problematical feature of the rooms. Also the size of the sweet spot in general was experienced to be too narrow by many interviewees.

Most respondents preferred rooms less reverberant, that is, reverberation should be set to a level where it does not make the room lively, but nor uncomfortable and unnatural. However, there were also opinions that room should sound more like living rooms when it comes to reverberation. Based on his own experience and literature, the author has noticed that especially within mastering engineers it is common to prefer a little more reverberant room.

When it comes to a stereo image, more accurate seemed to be the most preferred, thus engineer should be able to pinpoint exactly where different instruments are located in the stereo image. This, of course, does not coincide with the preference of wider listening spot. Envelopment in a sense of surrounding reflections was preferred in general, since it makes the room sound more natural. However, naturalness could not be done at the expense of smearing the stereo image and focus.

Frequency balance was experienced important and room should enable engineers to hear the full audio frequency range without any colorations. A great number of engineers also mentioned that a good room should enable an increase in sound pressure level (SPL) without added coloration to frequency balance.

A notable point from the survey was also the extensive use of near field monitors. Many engineers answered that they use at least 80 percent of the time near-field monitors. There were several reasons for that. First, they were experienced to reveal problems better and to give a better view on how music would sound in domestic environments. Second, the main monitors were said to sound "too good" by many engineers.

In their survey, Fazenda and Davies also asked from interviewees for a description, in their own words, upon how optimal control room should be. In the following, there are citations from [21, p. 8], where sound engineers describe the optimal control room:

- *"Clean, clear, detailed, strong..."*
- *"A neutral environment where you can make informed decisions about the recording you're making."*
- *"Balanced across the frequencies sure, you don't want it to be getting excessively bottom endy when you turn the speakers up."*
- *"I guess clarity, clarity is a big thing, I suspect a lot of reflections make confusing space."*
- *"A room with a flat response that doesn't mislead you, you want a large working area where there is very very little alteration in the sound you hear."*
- *"I really do like very accurate rooms, a good sounding room should be accurate with regards to image, I don't think a room should have any sort of imposed character"*.

3.5 Common specifications for an ideal control room

From the number of different design principles discussed in Chapter 3.1, and the variation in subjective opinions discussed in Chapter 3.4.3, it can easily be concluded, that there is no single answer how control room should be designed. There are different needs, such as, mastering, mixing, recording, music producing, and all of these may need different design approaches. For instance, the amount of money and space available, the user preference and the main genre to operate with may influence the design of the room. Despite all this, the author formed common goals which every professional control room should fulfill. In the following, there is a list of specifications for the optimal control room, which author has adopted from the previous research in control room acoustics.

1. Flat frequency response at least in the listening position (from 20 Hz to 20 kHz ±4 dB) [27, 52].

2. Frequency balanced reverberation time of 0.25 s [39, 58].

3. Proper distribution of room modes to produce accurate low frequency reproduction [58].

4. Initial time delay gap of 20 ms, that is, after the direct sound there should be no reflections over -15 dB during the first 20 ms.

Based on the literature review, a common agreement seems to be that the ideal room for critical listening has to provide the professionals possibility to be one step ahead of the consumers, that is, sound engineer can detect things that may become problematic on the best domestic systems [39]. On the other hand, control rooms have to be pleasing and creative environments where a mixing or mastering engineer can work for long periods of time. This means that acoustically too dry (almost anechoic) rooms are not acceptable. From time to time, debate rises if the control room properties should be more close to the domestic listening rooms, but all research based on subjective interviews with mixing engineers seems to conclude that majority of mixing engineers prefer much drier and precise listening environments than domestic listening rooms can provide [19, 39]. Of course, there are always exceptions, and as sound engineers are individuals it is quite hard to make any standardization for control room properties. This unfortunately makes the design of domestic hi-fi systems and rooms problematic as it is impossible to find the desirable properties of domestic systems that reproduce different music from the same premises. If a standardization would be made for control room design, it would be much easier to decide the properties for domestic hi-fi-systems.

3.6 Summary

In general, it seems that in the control room acoustics literature, more questions are asked than answers given. To better understand the perceptual relevance of acoustical events in control rooms, more subjective research and input from mixing engineers is needed. As Toole reminds in [56]: "When looking at the results of data gathered in 'scientific' circumstances, it is essential to think carefully before drawing conclusions about what may or may not be important in real-world situations".

4 Measurement and analysis methods

This chapter presents the measurement and analysis methods of this work. First, Chapter 4.1 introduces measurement arrangements in spatial impulse response measurements. After that, Chapter 4.2 presents spatiotemporal analysis in detail including directional analysis and visualization. Finally, the auralization is presented in Chapter 4.3 and the procedure of listening tests is revealed in Chapter 4.4.

4.1 Measurement arrangements

In this thesis, eight control rooms and four additional domestic rooms are investigated. Control rooms have been chosen considering different types of control room designs. Domestic rooms are examined for reference to enable comparison between professional control room and the room where music is commonly listened to as an entertaining purpose. Detailed information about the studied rooms are given in Table 2.

The goal of the spatial impulse response measurements is to get the whole three dimensional sound field of a room analyzed. Traditional monaural impulse response can only tell that reflections are coming in different time instants at different amplitudes, but it cannot specify from which direction individual reflections are coming to the receiver position.

Spatial impulse responses are measured using G.R.A.S 50VI 3D vector intensity probe. The probe consists of three pairs omni-directional microphone capsules facing to ±x, ±y and ±z directions. Three different sized spacers (2.5 cm, 5.0 cm and 10.0 cm) are used between microphone capsules to get more accurate localization for different frequency bands. A logarithmic sweep method is used for measuring individual impulse responses [18]. Logarithmic sweeps from 20 Hz to 24 kHz are recorded through Motu Ultralite Mk 3 audio interface to Macbook Pro computer using Reaper software with a sample rate of 192 kHz. The analysis is implemented with Matlab software. Figure 11 illustrates the signal path in the measurements.

All control rooms are measured using their own monitor loudspeakers as a sound source. In one of the control rooms also alternative monitors are used. For each speaker, A-weighted SPL was adjusted to 87 dB at a distance of 1 m using Sinus Tango Class 1 sound level meter. Figure 12 illustrates the receiver positions used in the measurements.

4.2 Spatiotemporal analysis

4.2.1 Directional analysis

The directional analysis (DIR) from spatial impulse responses is based on assumption that spatial impulse responses can be presented as a set of limited number of image-sources. The employed localization method estimates the 3-D location of the arriving sound field for each discrete audio sample with respect to the geometric center of the microphone array [54].

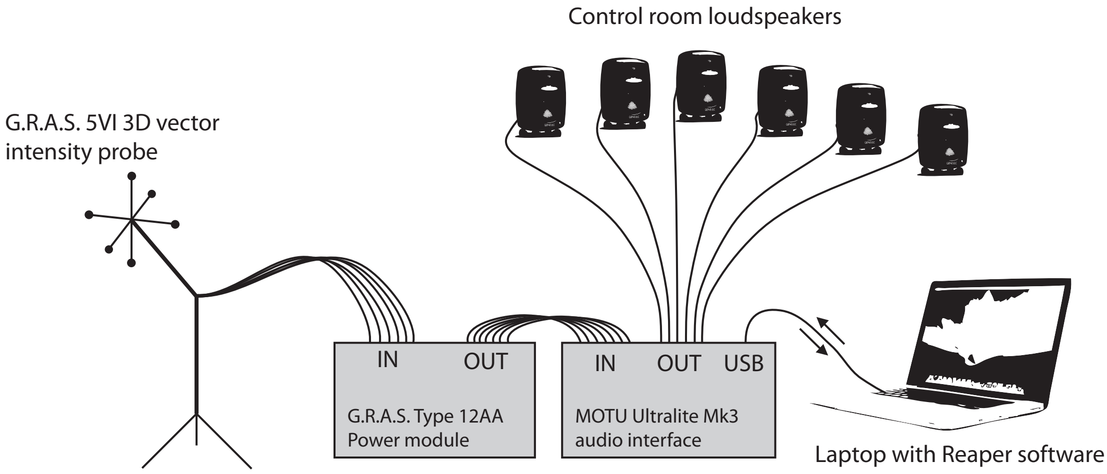


Figure 11: Measurement signal path. Logarithmic sweep signal is played with laptop through Reaper software, further to Motu Ultralite Mk3 audio interface from which it is fed to studio monitor loudspeakers. Room-loudspeaker response is simultaneously recorded with G.R.A.S. 5VI vector intensity probe via Motu Ultralite Mk3 audio interface to the computer. G.R.A.S. Type 12 AA Power module is used as a power supply for preamplifiers in the intensity probe. Power module applies 200V polarization voltage for each microphone capsule.

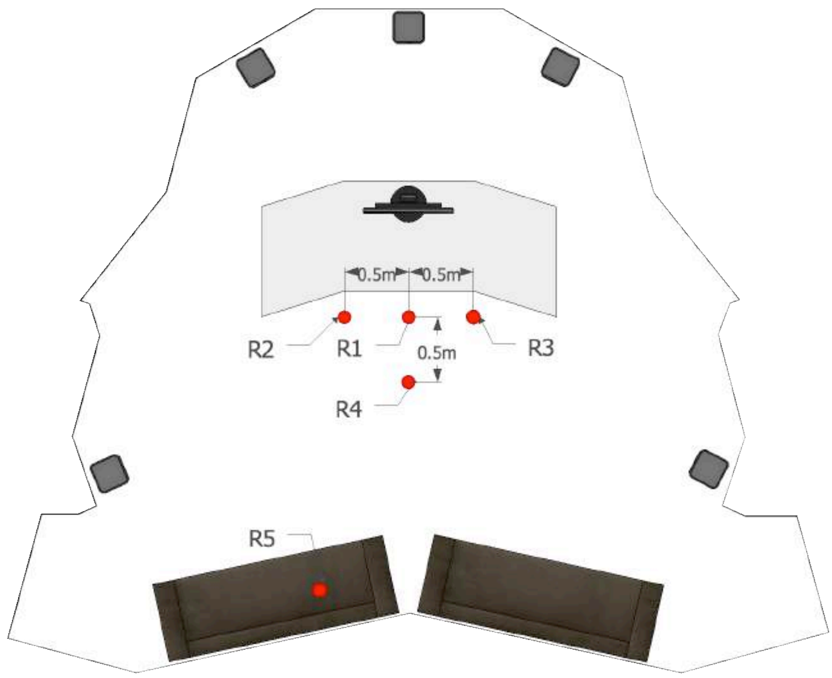


Figure 12: Receiver positions used in the measurements. R1 is the sweet spot, 0.2 m from the edge of the mixing table. Height of R1-R4 is 1.2 m and height of the R5 is 1.07 m.

If room impulse responses are given as $\mathbf{h}(t) = \{h_n(t)\}_{n=1}^N$, where N denotes the different microphone, the localization proceeds as follows. For each discrete time step, i.e., at every $\Delta t = 1/f_s$, where f_s is the sampling frequency, the room impulse responses are windowed with Hanning window. The window length is varied depending on the applied spacer. For 10.0 cm spacer, the window length is 0.69 ms, for the 5.0 cm spacer, 0.4 ms and for the 2.5 cm spacer, 0.25 ms, respectively. The window is centered at the sample of interest k. Next, generalized correlation method estimates the Time Difference of Arrival (TDOA) between each microphone pair. TDOA between microphones i and j can be written:

$$\hat{\tau}_{i,j}^{(k)} = \arg\max_{\tau}\{R_{i,j}^{(k)}(\tau)\}, \tag{11}$$

where $\arg\max_{\tau}\{\}$ stands for the argument of the maximum value, and

$$R_{i,j}^{(k)}(\tau) = \mathcal{F}^{-1}\{H_i^{(k)}(\omega)(H_j^{(k)}(\omega))^*\}, \tag{12}$$

where $\mathcal{F}^{-1}\{\}$ is inverse Fourier-transform, $H_i^{(k)}(\omega)$ frequency domain representation of windowed impulse response, and $()^*$ denotes complex conjugate. Next, each TDOA estimate is interpolated with the exponential fit (subscripts omitted for simplicity):

$$\hat{\tau} = \hat{\tau}^d + \delta, \tag{13}$$

where $\hat{\tau}^d$ is the original TDOA estimate from Eq. (11) and

$$\delta = \frac{(\log(R(\Delta t + 1)) - \log(R(\Delta t - 1)))}{4\log(R(\Delta t)) - 2\log(R(\Delta t - 1)) - 2\log(R(\Delta t + 1))}, \tag{14}$$

which is solved from exponential equations encountered when fitting the exponential function to original TDOA estimate [66].

The set of TDOA estimates is denoted with

$$\hat{\boldsymbol{\tau}}_k = [\hat{\tau}_{1,2}^{(k)}, \hat{\tau}_{1,3}^{(k)}, ..., \hat{\tau}_{N-1,N}^{(k)}]^T, \tag{15}$$

where N is the number of microphones, and the corresponding microphone position difference vector with

$$\boldsymbol{V} = [\boldsymbol{r}_1 - \boldsymbol{r}_2, \boldsymbol{r}_1 - \boldsymbol{r}_3, ..., \boldsymbol{r}_{N-1} - \boldsymbol{r}_N]^T, \tag{16}$$

where $\boldsymbol{r}_N$ is the position vector of the corresponding microphone. Solution for the direction vector is given as:

$$\hat{\boldsymbol{m}}_k = \boldsymbol{V}^+\hat{\boldsymbol{\tau}}_k, \tag{17}$$

where $()^+$ is Moore-Penrose pseudo-inverse. Moore-Penrose pseudo-inverse is used since the matrix encountered is not invertible. Direction of the arriving sound wave is given as $\hat{\boldsymbol{n}}_k = -\hat{\boldsymbol{m}}_k/||\hat{\boldsymbol{m}}_k||$.

DIR assumes that the sound field consists of plane waves and further that the microphone array is in the far field in relation to sound sources (loudspeakers and reflective room boundaries). A plane wave propagation model is assumed for the localization since an efficient estimator for the problem exists. After solving plane wave directions from the TDOA estimates, the arrival direction of every sample can be translated to azimuth and elevation angles $[\hat{\theta}, \hat{\phi}]$ via standard coordinate transformations.

After solving directions for each time step, DIR assigns a pressure value from an omnidirectional pressure impulse response for each of the locations. Ideally, the pressure value is obtained from an omnidirectional microphone capsule that is in the center of the microphone array. In this work, the pressure value is taken from the topmost microphone capsule. The approximation is made as the dimensions of the microphone array are small in comparison with the dimensions of the measured rooms, and thus the error is expected to be minor.

Now each spatial impulse response is presented by three values $[h_p(\Delta tk), \hat{\theta}(\Delta tk), \hat{\phi}(\Delta tk)]$ at each time moment Δtk. This data can be now used to plot spatiotemporal visualizations and to render auralizations. More detailed explanation of the localization is presented in [44, 54].

It must be noted that the window length limits the sensitivity of the direction estimation. This means that reflections are estimated fully independently if their separation in time is more than window length used in calculation (In this work 0.69 ms, 0.4 ms and 0.25 ms depending on the spacer). As another remark, DIR can be applied for arbitrary amount of microphones, and when the amount of microphones increases, the localization accuracy increases.

4.2.2 Visualization

To view spatial energy arriving in different time moments, spatiotemporal visualizations are used. First, polar response is calculated for each plane (lateral, median, transverse) from three values $[h_p(\Delta tk), \hat{\theta}(\Delta tk), \hat{\phi}(\Delta tk)]$ at each time moment Δtk. Polar response is derived as a cumulative sum of energy arriving from each direction in a considered plane during the specific time interval. In this work, visualizations are normalized so that the direction of maximum cumulative response, i.e., the direction of direct sound, is set to 0 dB. Time interval studied in this work was from direct sound to 50 ms after the direct sound, as it gives all the interesting information about the early reflections.

In spatiotemporal visualizations of this work, toroidal weighting is used for improving the energy separation between the lateral, median and transverse planes. Weighting functions reduce the effect of energy arriving from angles perpendicular to the analysis plane. Following toroidal weighting function is used for the lateral plane:

$$w_{lat}(\phi) = |\cos(\phi)|. \tag{18}$$

Similarly, the weighting is applied for other planes.

Figure 13 shows an example of spatiotemporal visualization over azimuth and elevation angles (lateral and median plane). The visualized responses are overlaid with the corresponding floor plan and cross section. Figure 13(a) illustrates the

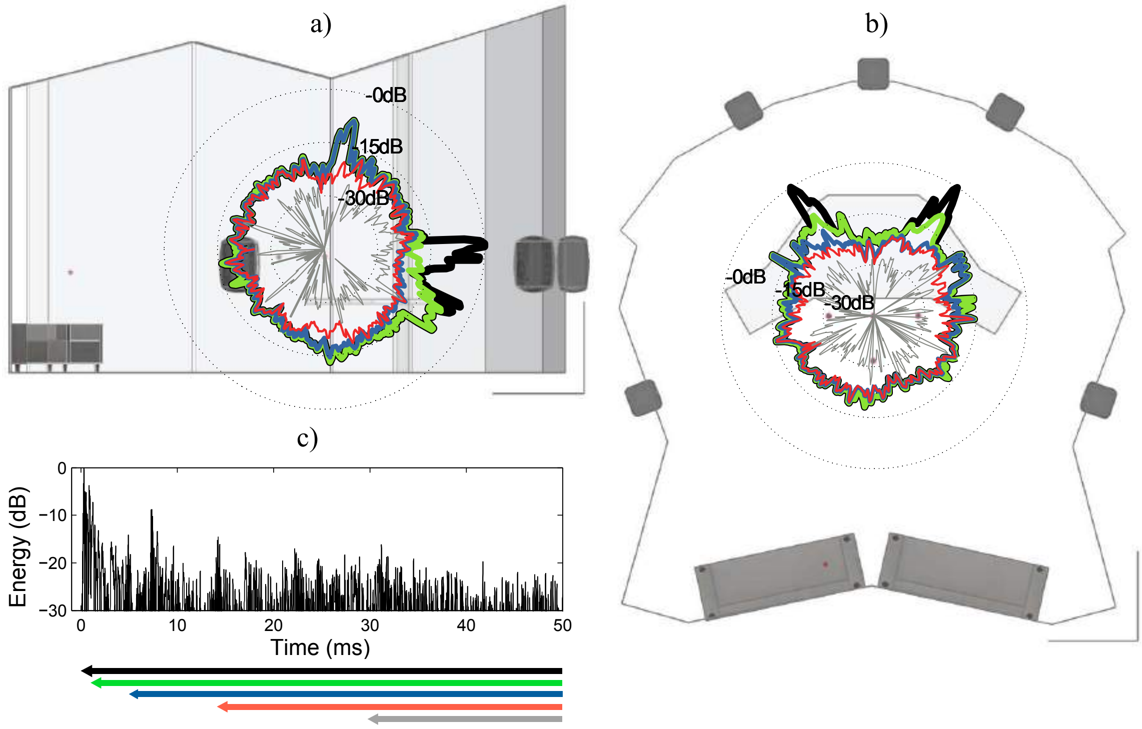


Figure 13: Example visualization of the spatiotemporal response at receiver position R1 in studio CR#6 from sources L and R, measured with 2.5 cm spacer. (a) Median plane. (b) Lateral plane. (c) Energy-time curve (ETC) in the same receiver position R1 which shows the applied time integration in (a) and (b). Thus, a black color corresponds the cumulative polar response from direct sound (DS) to 50 ms after DS, green from 1 ms after DS to 50 ms after DS, blue from 5 ms to 50 ms, red from 10 ms to 50 ms and gray from 30 ms to 50 ms after direct sound respectively. To reveal the dimensions of rooms, 1 m x 1 m scale is included in the bottom right corner of every spatiotemporal visualization as seen in both sub figures (a) and (b).

cumulative polar response in the median plane, and Fig. 13(b) in the lateral plane respectively. Figure 13(c) illustrates color correspondence to time window, thus, a black color corresponds the cumulative polar response from direct sound (DS) to 50 ms after DS, green from 1 ms after DS to 50 ms after DS, blue from 5 ms to 50 ms, red from 10 ms to 50 ms and gray from 30 ms to 50 ms after direct sound. The same color correspondences to time windows are used in all spatiotemporal visualizations within this thesis.

4.2.3 Automatic orientation calibration for the measurement probe

During the measurement procedure, it was noticed that adjusting the probe orientation precisely correct is practically impossible. In the preliminary visualizations, in certain studios, the direct sound was coming from slightly wrong direction compared with the measured loudspeaker positions. To compensate the possible user error occurring when adjusting the orientation of the measurement probe, and en-

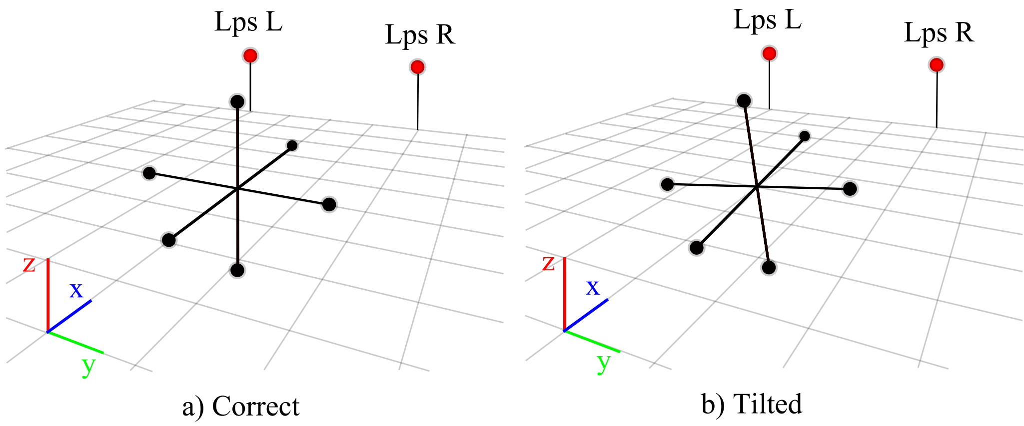


Figure 14: (a) illustrates the correct probe orientation and (b) situation when probe is tilted due to the user error.

suring the visualizations are correct, automatic orientation calibration was made in the analysis stage.

In automatic calibration, the loudspeaker positions are estimated from the measurements. After that, a plane which realizes the loudspeaker position coordinates is set via minimizing the sum of plane–loudspeaker distances in a least squares sense. As the loudspeakers are assumed to be approximately at the same height, the plane is expected to describe the orientation of the probe quite accurately. Finally, the rotation matrix [24, p. 190] is calculated to rotate the plane in x,y and z coordinates to correspond the situation in correct orientation, which is so that x-y plane of the probe is parallel to the floor and x-z plane is perpendicular to the difference vector between front stereo loudspeakers. The idea of rotation situation is illustrated in Fig. 14. Figure 14 (a) illustrates the correct probe orientation and Fig. 14 (b) situation when probe is tilted and correction is needed.

4.2.4 Reflection tracker GUI

To more accurately explore reflections, a graphical user interface (GUI) was implemented with Matlab. Reflection tracker GUI enables to explore the differences in results between different spacers, receiver positions and loudspeakers more quickly. In addition, the GUI enables to explore the room reflections more flexibly and within shorter time windows. Figure 15 illustrates the users interface. In the top left corner, there is drop-down menus to choose parameters. In the 'Studio' menu, user can choose which control room to explore. 'R' stands for different receiver positions from R1 to R5. From the 'Channels' menu, it can be chosen which channels to include to visualization. In Fig. 15, 1 and 2 stands for front stereo loudspeakers. In 'Probe' menu, there is three alternatives; small, medium and large, corresponding 2.5 cm, 5.0 cm and 10.0 cm spacers. 'Source' menu contains all source alternatives where user can choose between main and alternative monitor loudspeakers. A slider is specified to choose 0.5 ms window at each step and plot corresponding polar response in three different planes as seen in three sub figures in the bottom. Time

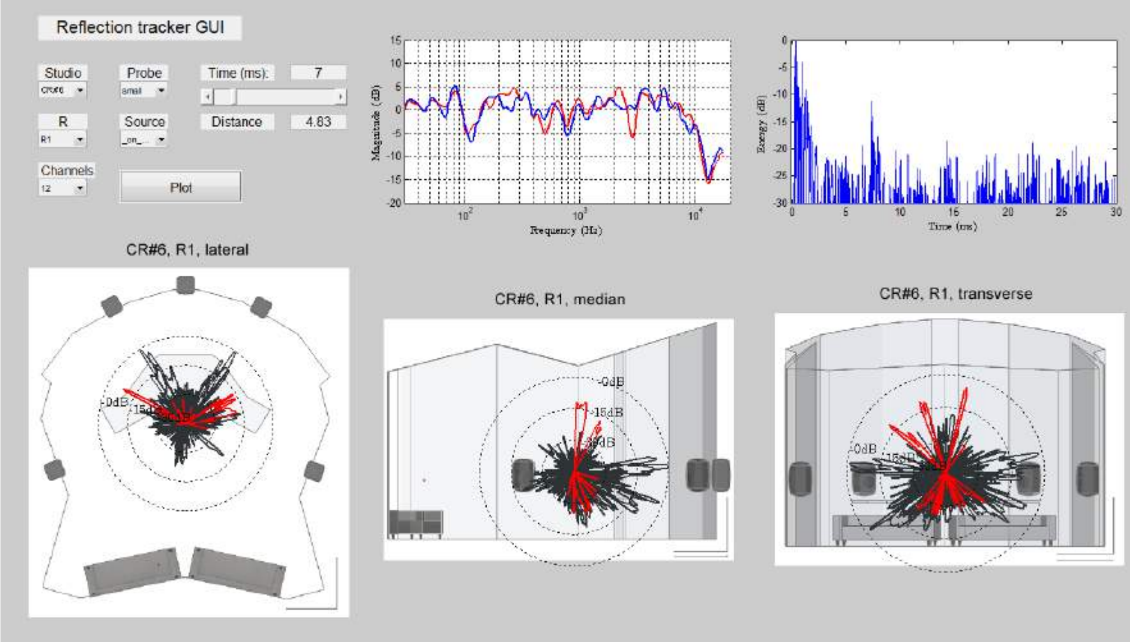


Figure 15: Reflection tracker GUI. Three subfigures in the bottom illustrates the polar response in specified 0.5 ms time window (red) and cumulative polar response before the time instant specified with the slider (dark gray). Center top sub figure illustrates magnitude response of channels specified and top right sub figure energy time curve, respectively.

related to the slider position is shown above the slider, 0 meaning the arrival time of the direct sound. There is also distance shown below the slider, which displays the distance sound travels within corresponding time.

In the spatiotemporal visualization shown in three sub figures located bottom of the GUI, red corresponds reflections coming within 0.5 ms time window before the time instant shown in the box above the slider. Dark gray illustrates the cumulative energy arrived before the time instant specified by the slider. Two subfigures in the top right corner illustrate the magnitude response and ETC with the parameters applied in the drop-down menus.

4.3 Auralization

To enable A/B comparison between different control rooms, sound field of each measured control room was auralized in an anechoic chamber. Listening system consists of 30 Genelec 8030A loudspeakers assembled in 3D-space. Figure 19 illustrates the loudspeaker positions used in auralization. Each music sample was convolved separately with the corresponding 30-channel SDM responses for left and right channel, respectively. In this work, each SDM-encoded sample was played back from the nearest loudspeaker with respect to the arrival angle. Also a virtual sound source positioning was concerned and tested with Vector Base Amplitude Panning (VBAP) technique [45], but playing each reflection through the nearest loudspeaker turned

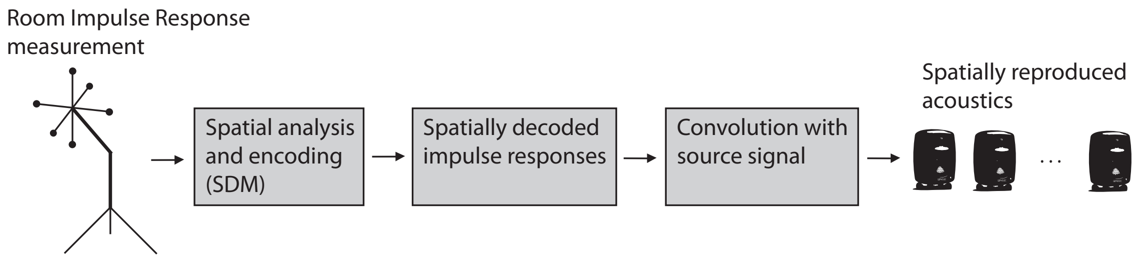


Figure 16: The block diagram illustrating steps in the auralization process.

out to give the best result for auralization. Figure 16 shows a block diagram about the auralization process.

Although the VBAP should give theoretically a bit more precise sound source positioning, it turned out that panning smeared the perceived sound and made it unnatural compared to the direct sound source positioning. As there are 30 loudspeakers and they are positioned to cover all directions of reflections, possible errors in the arrival angle of the synthesized samples are not very large. The directions of the most significant reflections were concerned when positioning loudspeakers.

Finally, the loudness of the auralizations was normalized. A-weighted SPL average was measured over 30 seconds with a reference sound sample (*Within Temptation - Faster* 46 s - 76 s) for each control room. After that, overall gains were adjusted to the same. These calibrations were also verified by listening to the samples. It must be noted that if loudness levels are matched, human cannot perceive distance in anechoic room [40], thus possible longer distance between listener and a loudspeaker in anechoic room, compared to the corresponding distance in measured control room is not a problem when rendering auralizations.

4.4 Listening tests

The aim of the listening tests was to get input from professional mixing and mastering engineers concerning their needs and preference in control room acoustics. With this preference data, it would be possible to discuss which kind of spatiotemporal response would be optimal. It is also tried to find out thresholds for deficiencies such as frequency response colorations and excessive reverberation. Final aim was to benchmark the success of the auralization methods.

4.4.1 Listening room setup

Listening tests were conducted in an anechoic chamber, located in Aalto University's Department of Signal Processing and Acoustics. Inner dimensions of the chamber are 4.3 m $\times$ 4.3 m $\times$ 4.3 m and it is asymptotically anechoic on frequencies above 100 Hz. Loudspeaker setup was the same as described in Chapter 4.3. Figure 19 a) illustrates loudspeaker positions as a wire-frame model and Fig. 19 b) is picture of the anechoic chamber. Red ellipses denote the left and right channels at $\pm$ 30 degrees in lateral plane in both subfigures 19 a) and 19 b). Auralizations were played back with Max/MSP, and subjects controlled the user interface with an iPad.

4.4.2 Test subjects and control rooms

In total 15 subjects participated in the listening test, out of which 12 are recording/mixing engineers and 3 of them are mastering engineers. Ten of the subjects are professionals and five are music technology students but all of them had at least 5 years experience in mixing or mastering. Ten of the subjects in the listening tests are the principal users or the owners of the control rooms included in this study. Detailed list of the participants can be found in the Table 1.

Only professional mixing or mastering rooms were included in the listening tests. Total of eight different rooms were included, and one room also with alternative loudspeakers. Thus, total of nine control room-loudspeaker combinations were included in listening tests (Rooms 1 - 9 in Table 2). Five of the rooms were control/mix rooms and three of them were mastering rooms. Only the sweet spot was studied in the listening tests. Detailed list of the control rooms can be found in Table 2

Table 1: List of subjects in the listening test. +15 equals 15 year or more.

ID	Main genres	Experience [y]
MI 1	jazz/folk/classic	10
MI 2	pop/rock	5
MI 3	pop/rock/jazz	5
MI 4	pop/rock	10
MI 5	rock	5
MI 6	jazz/folk/classic	10
MI 7	electronic	5
MI 8	pop/rock	5
MI 9	acoustic/pop/rock	+15
MI 10	acoustic/pop/rock	+15
MI 11	pop/rock	+15
MI 12	jazz/classic	5
MA 1	pop/rock/electronic	10
MA 2	pop/rock	+15
MA 3	rap/rock	+15

4.4.3 Listening test procedure

Before the listening test, hearing of the participants was tested with standard audiometer. Actual listening test consisted of three parts. In the first part, participants got used to all nine control rooms by listening five different songs in all rooms. Songs in the familiarization step included:

Table 2: Control rooms and loudspeakers included in the listening tests. T_{60} is calculated from wide band decay from -5 dB to -25 dB. CR# is the number of the room loudspeaker combination, LPS is loudspeaker, A is floor area, V is the volume and d is the distance between loudspeakers and listening spot. Rest of the abbreviations are given in the end of the table. Note that CR#2 utilizes the same room as CR#3, but the difference is that CR#2 uses near field monitors, and CR#3 flush mounted monitor loudspeakers.

CR#	LPS	Purpose	A [m^2]	V [m^3]	d[m]	T_{60} [s]	f_c [Hz]
1	3W-p	MA	18.7	57.5	1.60	0.38	163
2	2W-a	C/MI	26.9	73.5	1.83	0.23	111
3	3W-a	C/MI	26.9	73.5	3.22	0.24	114
4	3W-a	MA	22.9	56.2	2.60	0.20	119
5	2W-a	C/MI/T	22.5	58.1	2.42	0.26	133
6	3W-a	C/MI/T	28.9	96.0	2.57	0.39	127
7	2W-a	C/MI	16.3	38.7	2.12	0.17	129
8	3W-a	C/MI/T	29.0	78.0	2.85	0.32	128
9	3W-p	MA	21.0	66.5	1.95	0.29	124
10	3W-a	HT	21.0	52.7	2.20	0.32	156
11	2W-p	PR/MI	15.3	42.8	0.90	0.22	143
12	2W-a	L	17.3	46.1	2.80	0.54	216
13	2W-p	L	42.4	105.8	3.60	0.41	132

a: Active loudspeaker, p: Passive loudspeaker, 2W: 2-way loudspeaker, 3W: 3-way loudspeaker

C: Control room, MI: Mixing room, T: Teaching room, MA: Mastering room, PR: Production room, L: Living room, HT: Home theater

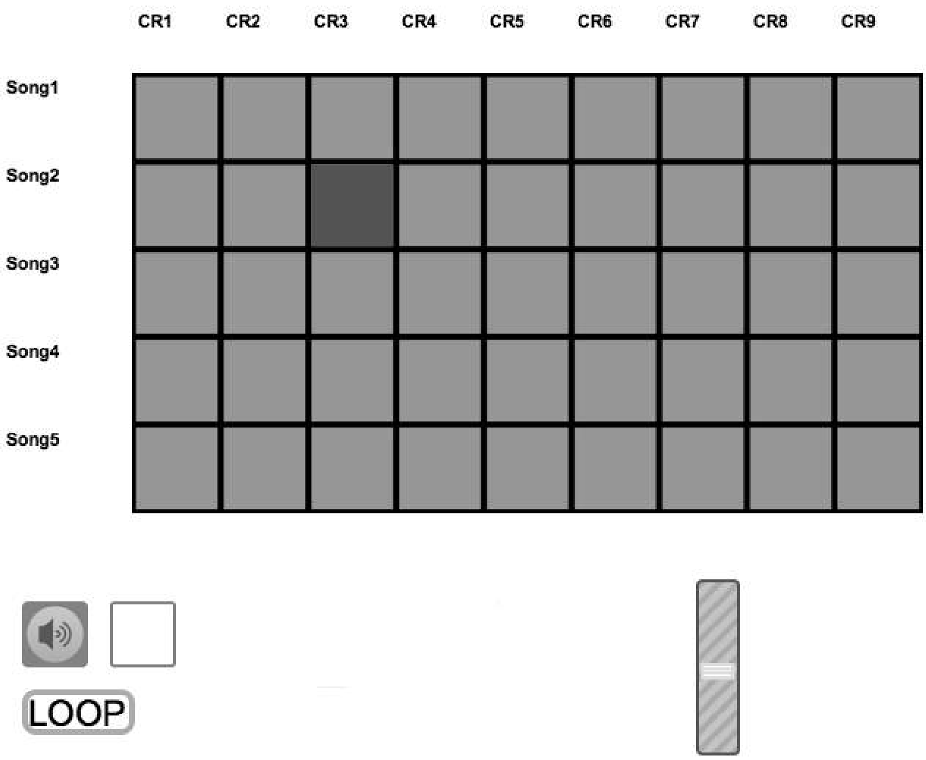


Figure 17: Max/MSP user interface used in both familiarization part and the interview part of the listening test.

- Song 1: *Shawn Colvin - A Matter of Minutes (60-90 s)*
- Song 2: *Within Temptation - Faster (64-94 s)*
- Song 3: *Jamiroquai - Cosmic Girl (20-50 s)*
- Song 4: *HIM - Love Without Tears (40-70 s)*
- Song 5: *Dick Oatts - Raised Nine Ball (260-290 s)*

Songs were chosen to have a combination of female voice, dry percussive elements and broad, dense spectra. As Toole states, the most useful recordings to reveal spectral or timbral differences in audio are concluded to have broad, dense spectra and percussive elements [56, p. 446]. Multiple comparison user interface was implemented with Max/MSP matrix and subjects controlled it with an Ipad. Figure 17 illustrates the Max/MSP matrix used in familiarization as well as in the interview part of the listening test.

Second part of the test was preference rating, which was carried out in a pair comparison manner. Thus, participants compared two control rooms at a time, and they were asked to choose the one they think better suit their personal needs in mixing or mastering. Participants were also advised to think about arguments for their decisions. Subjects were able to adjust the master level during the test. Sound samples in pair comparison test included:

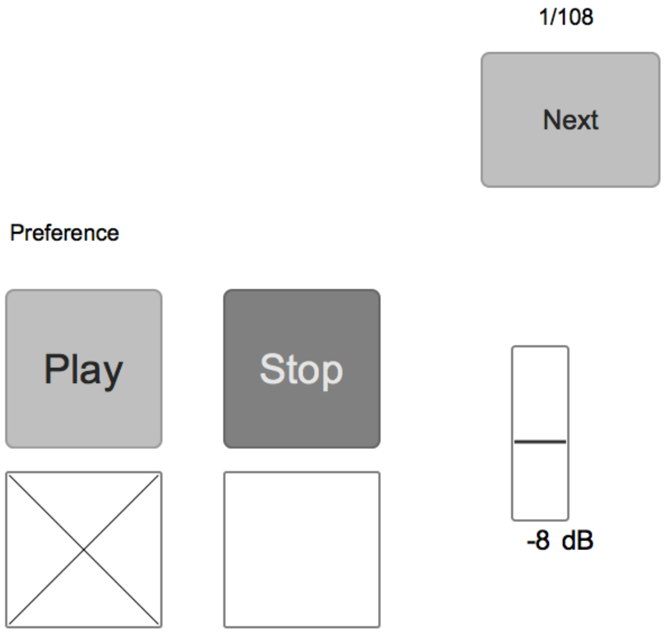
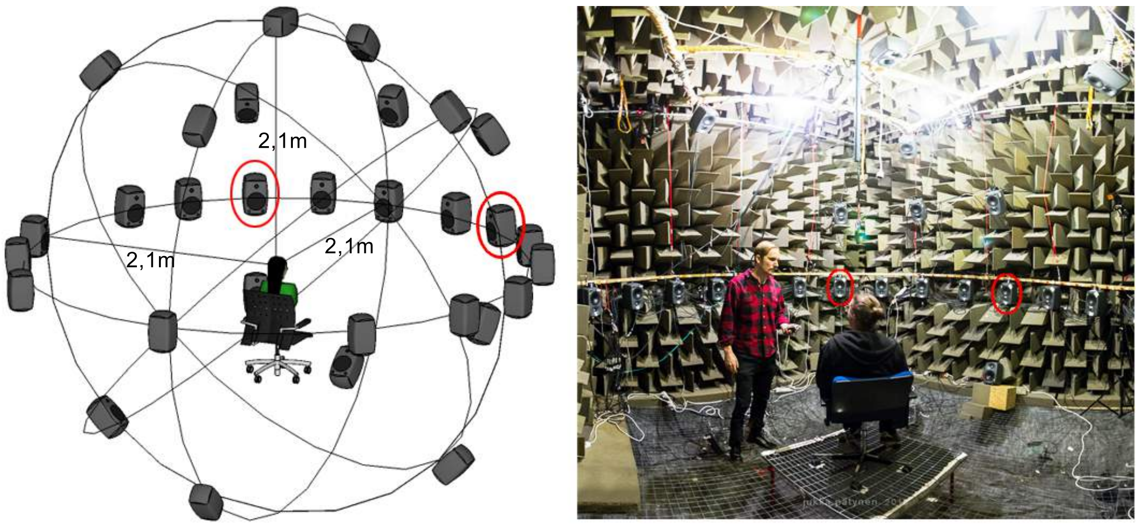


Figure 18: Max/MSP user interface used in the pair comparison preference test.

- Sample 1: *Shawn Colvin - Matter of Minutes (60-72 s)*
- Sample 2: *Within Temptation - Faster (64-78 s)*
- Sample 3: *Jamiroquai - Cosmic Girl (20-32 s)*

All nine control rooms were covered with all three sound samples, thus total of 108 pairs were included in the pair comparison test. Pair comparison test user interface is illustrated in Fig. 18.

Third part of the test was an interview where subjects were asked for their arguments for preference. After that, control rooms were listened one by one, and participants were asked to describe each room with their own words. In an interview, it was also asked if subjects can find their own control room if it was included in the study. Finally, a brief survey was done about the history and experience of the participants.



a) b)

Figure 19: Listening setup used in the listening tests. Loudspeakers corresponding $\pm 30°$ in lateral plane, i.e., ITU-R standard [27] left and right stereo loudspeakers are marked with red ellipses in both subfigures a) and b).

5 Results

This chapter presents the results of this work. First, measured control rooms are analyzed on the basis of objective spatiotemporal visualizations in Chapter 5.1. Second, the results from the listening tests are revealed in Chapter 5.2, including the preference ratings and the summary of the interviews.

Measured rooms are listed in a Table 2, which shows that rooms differ quite much from each other when it comes to a room volume, surface area or a reverberation time. The surface area of control/mix rooms and mastering rooms varies from 16 to 29 square meters and average T_{60} from 0.17 s to 0.39 s respectively. Calculated f_c varies approximately between 110 and 130 Hz in these rooms. However, it must be kept in mind that in small rooms, calculated f_c values are usually lower than actual values.

Comprehensive measurement results can be found from Appendix A, which includes spatiotemporal visualization, ETC, magnitude response, and CSD of each measured room at receiver position R1, i.e., the sweet spot. In addition, magnitude response comparison figures can be found from Appendix B.

5.1 Spatial impulse response measurements

From the spatiotemporal visualizations presented in Chapter 4.2.2, room reflections can be tracked both spatially and temporally. The cumulative amount of early energy can be clearly observed from visualizations. This chapter discusses about possible causes behind acoustic reflections that can be tracked from the visualizations.

5.1.1 Lateral plane

Figure 20 illustrates spatiotemporal visualization of six measured control rooms in lateral plane. The ratio between direct sound and early reflections can be easily observed from Fig. 20. In 20 a) and f) the direct sound to early reflections ratio is significantly higher than in 20 b), c), d) and e). This happens as the distance between the listening position and loudspeakers is smaller in control rooms a) and f), but also because there is obstacles between loudspeakers and the listening position in rest of the control rooms. These obstacles, such as alternative monitors, racks, and large tables cause reflections that decrease the ratio between the direct sound and early reflections. Diffusivity of the early reflections can also be easily compared from the Fig. 20. It can be seen that in 20 a) and e) cumulative polar response is much more rectangular than in 20 b), c), d) and f).

Also the symmetry of the early reflections can be observed from the spatiotemporal visualization in lateral plane. In professional control rooms, symmetry should be high, but there is still a slight asymmetry in the early reflections as can be seen in Fig. 20 d). The early energy coming from left side of in 20 b) is probably due to a calculation error. At least no physical reason for these reflections could be derived.

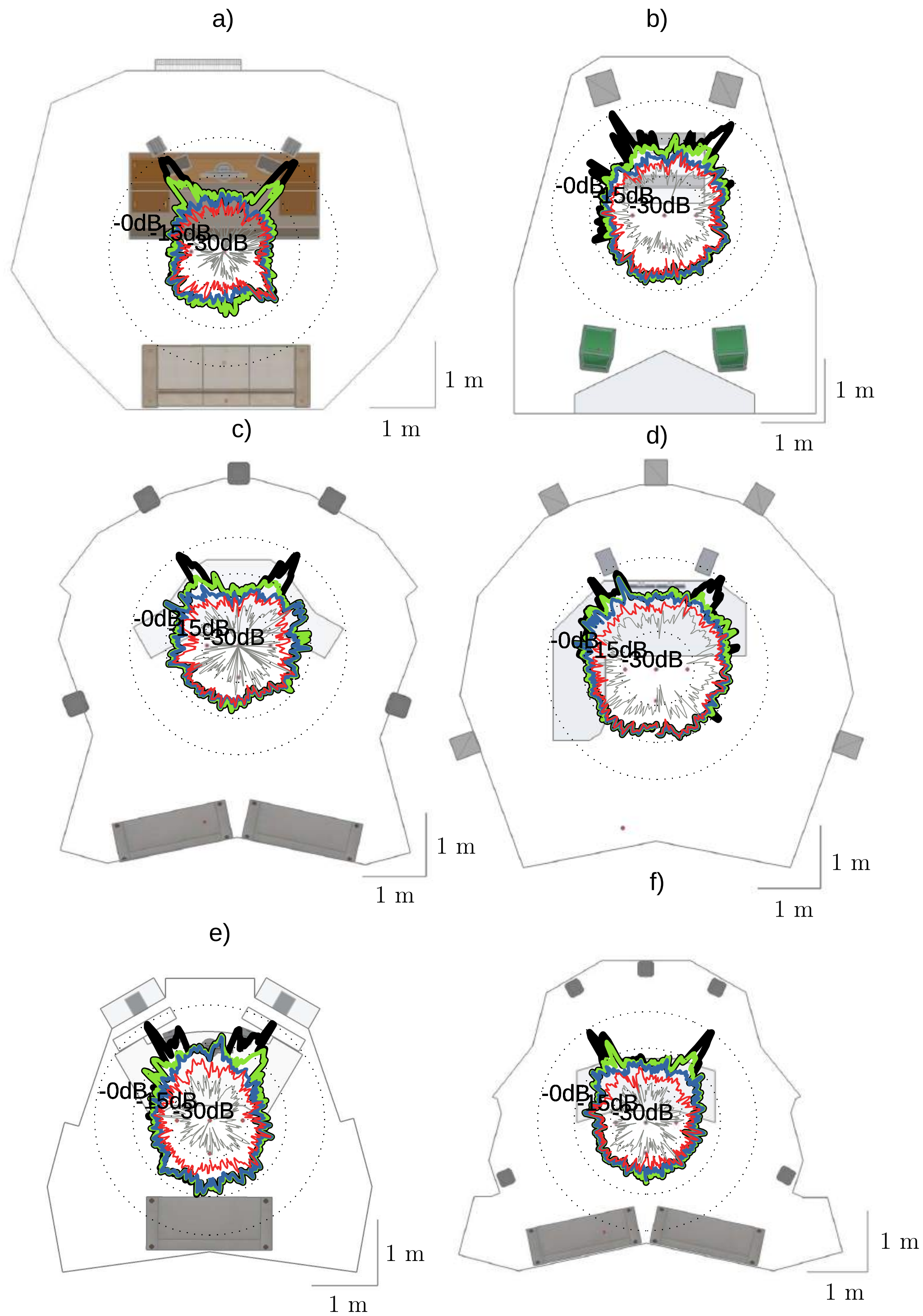


Figure 20: Cumulative polar response of six measured control rooms in lateral plane. a) CR#2, b) CR#9, c) CR#6, d) CR#8, e) CR#7 and f) CR#5.

5.1.2 Median plane

Figure 21 illustrates spatiotemporal visualization of six measured control rooms in median plane (in same order as in Fig. 20). It can be seen a mixing console/table reflection in control rooms (a), (c), (e) and (f). In Fig. 21 (a), the console reflection is almost as strong as the direct sound being only a couple of dB lower compared to the direct sound. This kind of effect can shift the sound image towards the direction of the table reflection. Ceiling reflection can be seen in Fig. 21 (b), (c), (d) and (f) being the strongest in (c). From the Fig. 21 it is evident that the table and its tilting has a major effect to the cumulative energy arriving obliquely from below. The amount of reflecting energy from the table depends also decisively on the positioning of monitor loudspeakers. Thus, to minimize the table reflection, attention should be paid to the loudspeaker and table positions in a control room.

Comparison of the spatiotemporal illustrations in appendix A shows clearly the effect of the mixing table. As an example, mastering rooms CR#1 and CR#9 (Figs. 29 and 37) has both relatively narrow and tilted table, thus the reflections from obliquely below are much lower than for example in rooms CR#2 and CR#6 (Figs. 30 and 34), which utilizes such a large mixing table.

5.1.3 Different receiver positions

Figure 22 illustrates spatiotemporal visualizations of CR#5 in three different receiver positions. Figure 22 a) and b) correspond receiver position R1 in lateral and median plane, c) and d) receiver position R4 in lateral and median plane, and e) and f) R5, respectively. It can be seen that the amount of the direct sound decreases when the distance between loudspeakers and the listening position increases. Also the back wall reflection can be easily seen from the Fig. 22 e) and f). Looking at the median plane, there is slight variation in the angle of the table reflections between 22 b) and d) but when moving all the way to the back of the room, are table reflections disappeared, or integrated to the direct sound, because of the different angle between the table and the listening position.

5.1.4 Control rooms and living rooms

Figure 23 shows spatiotemporal visualization of two control rooms (a and b) and two domestic living rooms (c and d), where a) is CR#9, b) is CR#5, c) is CR#12 and d) is CR#13. It can be seen that in control rooms, the shape of the polar response is rather circular if the direct sound is neglected. This means that the early sound field is diffuse in control rooms. This can be seen in Fig. 23 (a) and (b). Contrarily in domestic living rooms, polar responses are more irregular, which tells about more unevenly distributed early reflections. It can be seen in Fig. 23 (c) that the direct sound is hardly distinguishable from the reflections, which is logical as the listening spot is quite far from the loudspeakers, and T_{60} is as high as 0.54 seconds. It can be also seen in Fig. 23 (c) that back wall reflections are almost as loud as the direct sound, which is due to the hard back wall just behind the listening position.

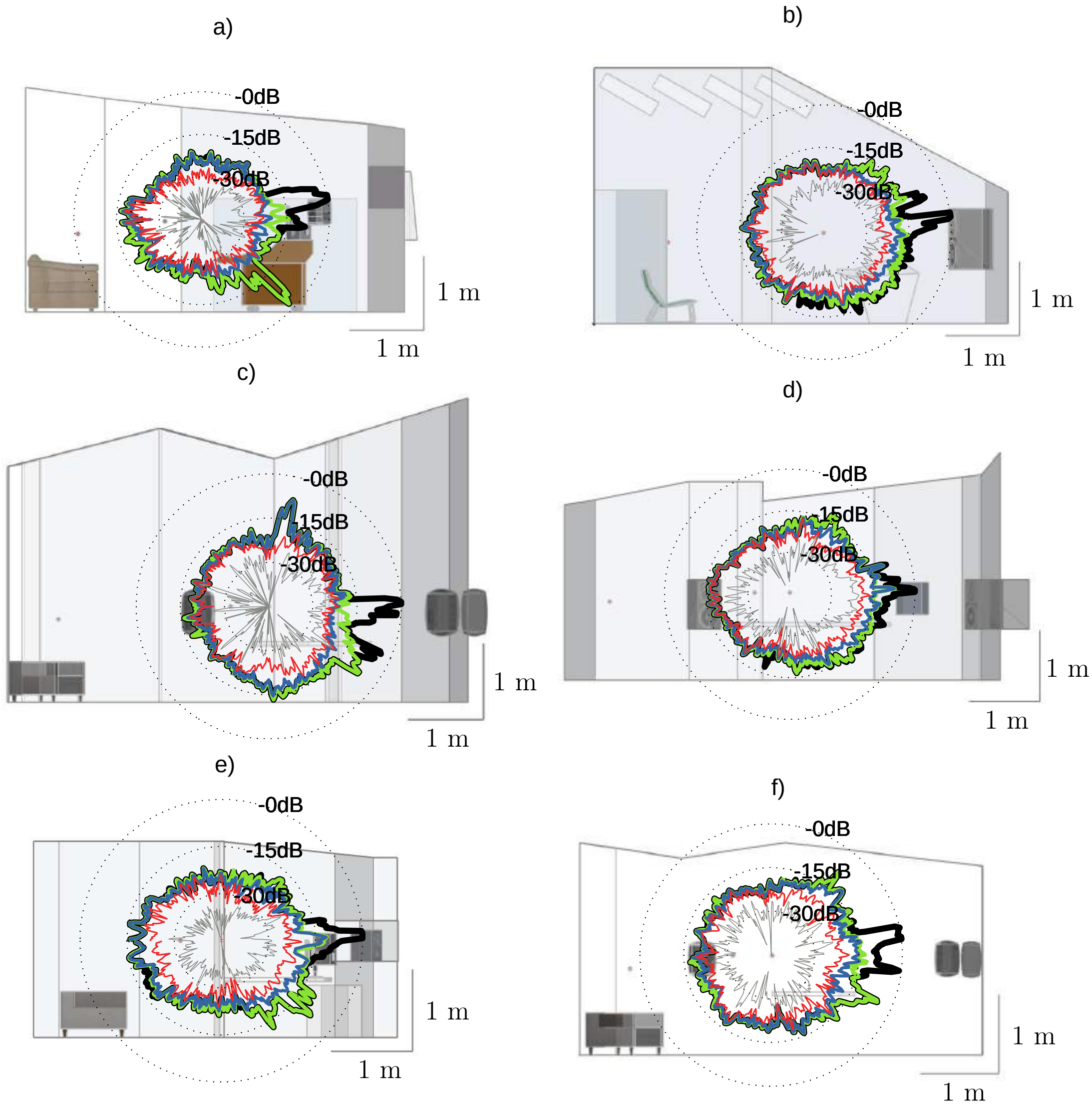


Figure 21: Cumulative polar response of six measured control rooms in median plane. a) CR#2, b) CR#9, c) CR#6, d) CR#8, e) CR#7 and f) CR#5.

Taking a closer look at the Fig. 23 d), it can be seen that the early sound field is quite irregularly divided. This is quite logical when observing the geometry of the room. There is an opening to another room on the right side of the listening position, which seems to affect to the amount of reflected energy from that direction. On the contrast, left side has a wall, and more energy is coming from that direction as can be seen in Fig. 23 d). Same kind of phenomenon can be seen in also in CR#12 illustrated in Fig. 23 c). This kind of irregularities in geometry can affect to the stereo balance and cause image shifts towards to the direction of more early energy.

At first glance, the spatiotemporal visualization of CR#12 (Fig. 23 c) might seem illogical when compared to the one of CR#13 (Fig. 23 d). Although the

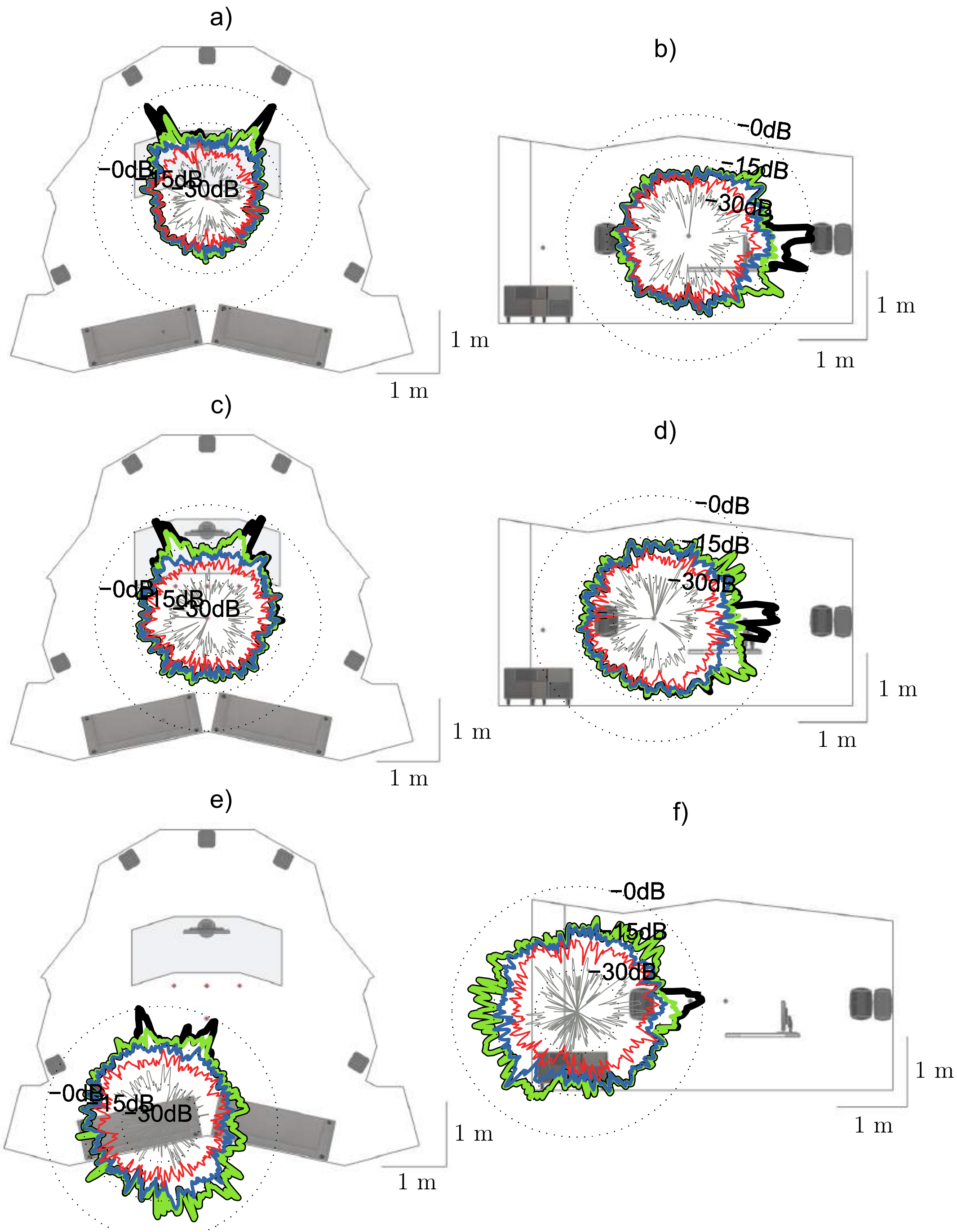


Figure 22: Cumulative polar responses of CR#5 in lateral and median plane at different receiver positions. a) R1 in lateral plane, b) R1 in median plane, c) R4 in lateral plane, d) R4 in median plane, e) R4 in lateral plane and f) R4 in median plane.

reverberation time difference is only 0.13 s, T_{60} being 0.54 s in CR#12 and 0.41 s in CR#13, and the loudspeakers are more far in CR#13, it seems that CR#12 has much more early reflected energy compared to the direct sound. This can be explained with the directivity of the loudspeakers in CR#13. The loudspeakers of CR#13 are electrostatic panels which are extremely directive, thus the direct sound is dominating in the spatiotemporal visualization of CR#13 (Fig. 23 d).

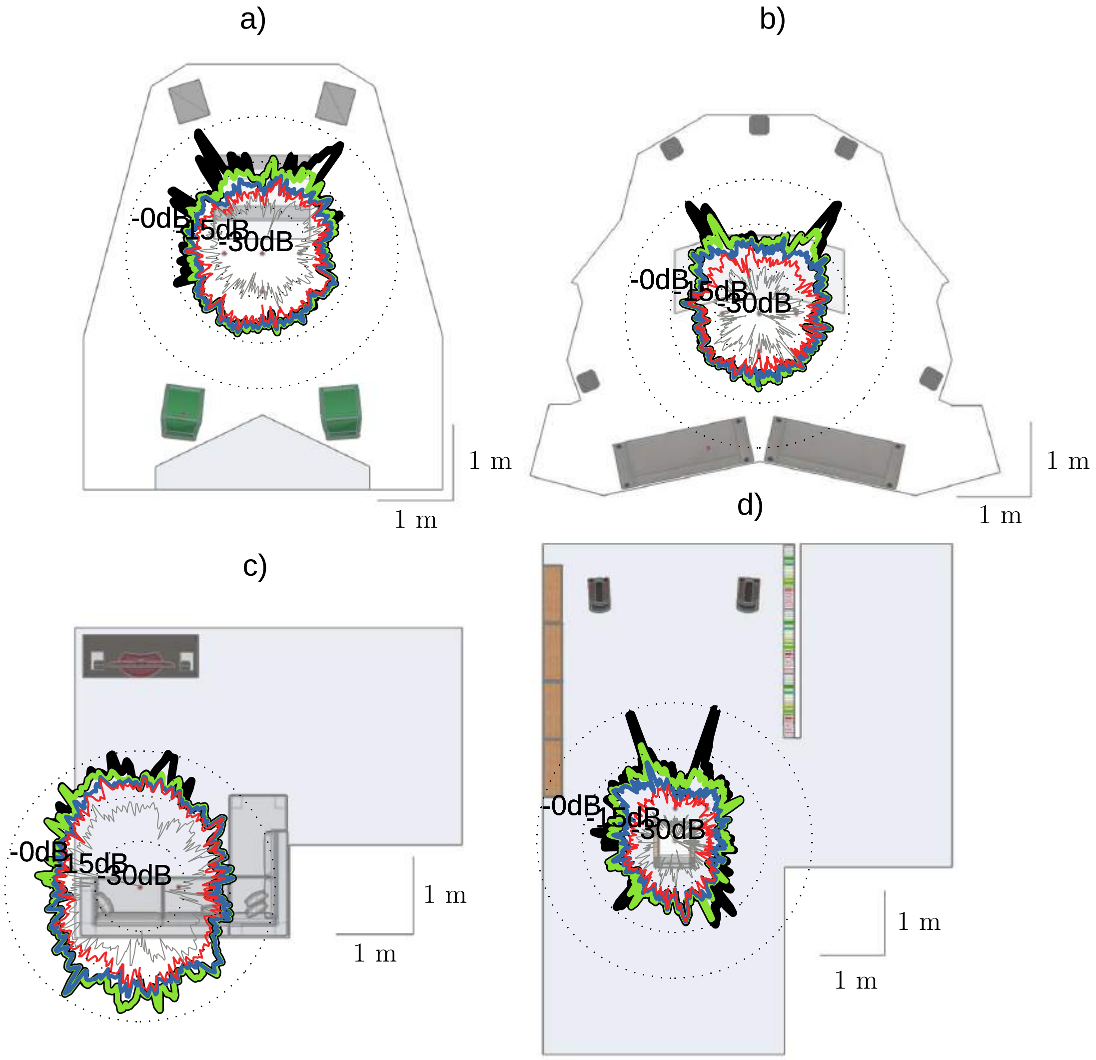


Figure 23: Cumulative polar response of two control rooms (a and b) and two domestic living rooms (c and d) in lateral plane. a) CR#9, b) CR#5, c) CR#12 and d) CR#13.

5.1.5 Shortcomings of the energy-time curve

In many measured rooms, there can be noticed individual spikes in the ETC plot, as can be seen in Appendix A. These spikes could be easily interpreted to be individual reflections, but after further investigation with reflection tracker GUI, it turned out that the energy is coming from many directions at that time instant. Figure 24 shows a distinguishable spike in ETC at 23 - 24 ms in CR#9 at the receiver position R1, but an investigation using the reflection tracker GUI reveals that sound energy

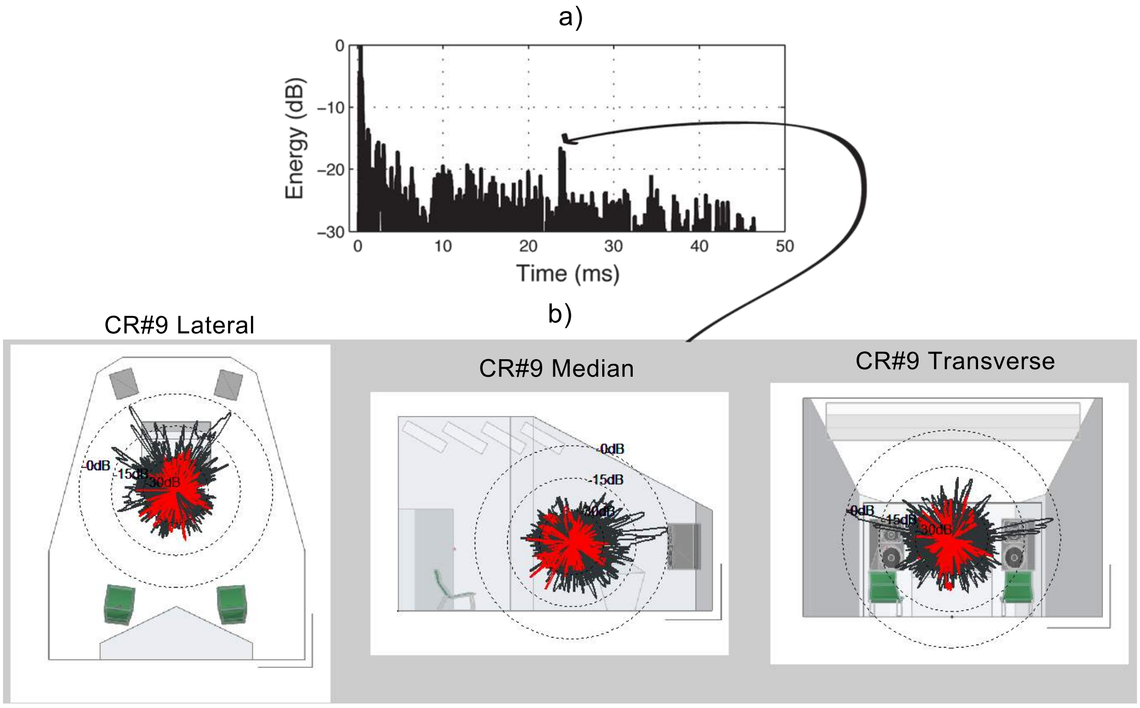


Figure 24: a) ETC of CR#9, b) cumulative polar response of CR#9 visualized in lateral, median and transverse planes. In b), red color corresponds the energy arrived during 23 - 24 ms after the direct sound. Gray is a cumulative polar response from direct sound to 24 ms after the direct sound, respectively.

is coming almost from all directions at that time. Phenomenon is probably related to room dimensions. As the listening position (R1) is somewhat in the middle of the room, focusing of reflections is a possible reason for the arriving energy spike. When reflected energy is arriving from many directions, it is not heard as disturbing than if it was an individual reflection from one direction. This proves that ETC does not always tell the truth about reflections and their perceptual levels, because it cannot tell whether the energy spike is due to the multiple focused reflections or an individual reflection from one direction.

If the success of the design goal of having no reflections over -15 dB compared to the direct sound during the first 20 ms is observed from the ETC plot, the perceptual relevance is not shown. That is, a single energy concentration in the ETC does not necessarily correspond to any disturbing or even audible reflection.

5.2 Listening tests

This chapter presents the results of the listening tests. First, preference rating is analyzed and reasons for the preference are discussed. Second, summary of the interviews is shown as translations from Finnish transcriptions.

5.2.1 Preference

Figures 25, 26 and 27 presents preference rating charts assembled on the basis of a pair comparison test. X-axis is the control room and Y-axis is the subject. Colors and a red number both denote how many times a subject has preferred certain control room. Colors correspond the number in a way that white is 0 and black is 8. Above, there is preference averages over mixing and mastering engineers for each control room.

It can be seen from the Figs. 25, 26 and 27, that preference rating varies significantly between subjects and also between sound samples. With Sample 3 it seems that mixing engineers have the most consistent preference over control rooms, as the most preferred and the least preferred rooms can be clearly separated in the Fig. 27. Among mixing engineers, CR#2, CR#4, CR#5 and CR#7 were the most preferred, and CR#6, CR#8 and CR#9 were disliked quite often. With Sample 1, there is a parallel behavior in preference to Sample 3, but the consensus is not as clear. The most random preference rating seems to be with Sample 2, as the rating varies a lot between subjects. It also seems that in the case of Sample 2, it has been hard for subjects to find the most preferred control room.

One reason for the variation in preference rating between sound samples is probably the fact that sound engineers were differently familiar with the samples. Majority of engineers reported during the interview, that the preference decisions were easiest with Sample 3 and especially Sample 2 was problematic song or genre for them. This is in line with the preference rating charts in Figs. 25, 26 and 27 as the amount of gray color indicates about uncertain preference within control rooms. Other reason can be the different amount of transient components between samples, and the amount of added reverberation to the samples. Many sound engineer reported the room reverberation being important factor for the preference. This is logical to the result that Sample 3 gave the most consistent preference, because it is a quite dry song with lots of transients, which makes it easy to perceive the room reverberation. With Sample 1 and Sample 2, the added reverberation undoubtedly blends to the control room reverberation, which can make the preference rating harder with the argument of control room reverberation.

Mastering engineers MA2 and MA3 preferred CR#1, CR#6 and CR#8 most. These rooms have longer reverberation times (0.38 s, 0.39 s and 0.32 s), than the rooms that mixing engineer preferred. These two mastering engineers gave consistent preference ratings with different sound samples. Mastering engineer MA1 preferred mostly CR#9 with Sample 3 and Sample 1, and CR#2 and CR#7 with Sample 2. It must be noted, that only three mastering engineers were included in the listening tests, thus no statistically reliable conclusions can be made for the mastering engineers as an occupation.

5.2.2 Interview

Following is a summary of interviews, which were conducted immediately after the preference test. The answers of 14 (Finnish speaking) subjects are presented as translations from Finnish transcriptions and the answers of one (English speaking)

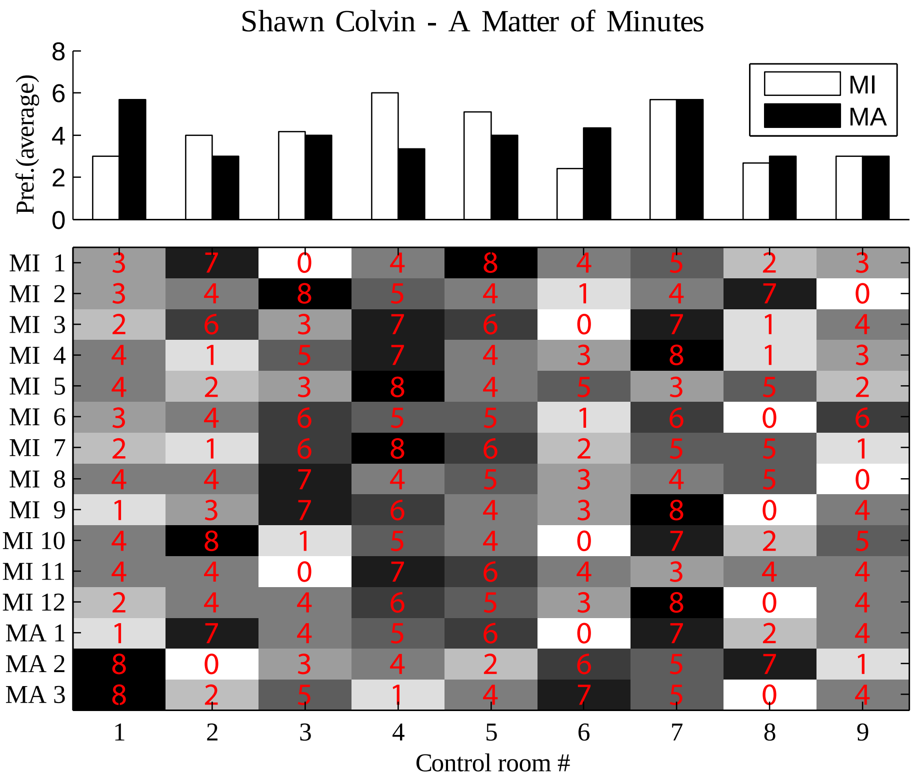


Figure 25: Preference rating chart on the basis of pair comparison test for the Sample 1 (Shawn Colvin - A matter of Minutes). MI: mixing engineer, MA: mastering engineer.

Table 3: The most common attributes that the subjects used in the preference ratings according to the interviews.

Width, accuracy, or stability of the stereo image (×8)
Reverberation, reverberance (×7)
Frequency balance (×7)
Localization: direction of distance of sources (×6)
Low frequencies: clarity, compactness, or booming (×4)
Clarity: high, low, or wide band (×3)

subject are presented as direct quotations. In the first part, two questions deal with the arguments for the preference rating and how realistic the auralization was experienced. In the second part, there is a summary of comments related to each control room.

Question 1: Describe the attributes which you find the most important in critical listening environment, thus which were your main arguments for your preference decisions?

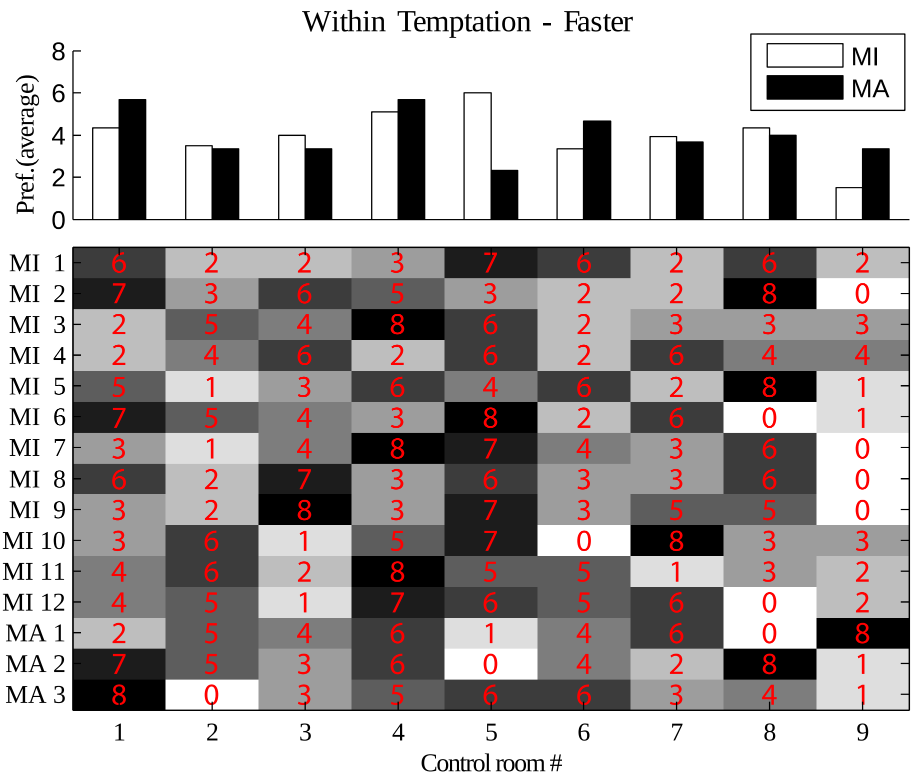


Figure 26: Preference rating chart on the basis of pair comparison test for the Sample 2 (Within Temptation - Faster). MI: mixing engineer, MA: mastering engineer.

MI 1:
"If the sound is not coming from the elevation of the loudspeaker and also not from the front, it makes me very confused. When mixing, you should listen to the acoustics that has been recorded, then it would be nice if the sound produced from the loudspeakers is free of room reverberance. If I would have chosen the preference according to which is more entertaining, I would have probably chosen more reverberant rooms. Frequency balance was more, kind of a, taste question, I mean that if you are used to work in a certain place, you adapt to the magnitude response and then can guess how the recording will sound in other places. What comes to frequency balance, there wasn't any (ed. rooms) with a poor balance, but I noticed comb filtering in a couple of rooms."

MI 2:
"I paid attention to some kind of compactness and firmness of the sound, especially at bottom-end (ed. low frequencies). I did not like the ones (ed. control rooms) that had significant amount of reverberation. On the other hand, sometimes I felt more surrounding room more comfortable. In a couple of rooms I felt that the stereo image is almost too narrow. In some rooms, I noticed disturbing boost at low frequencies and that was a reason to choose the other. (ed. in the pair comparison)"

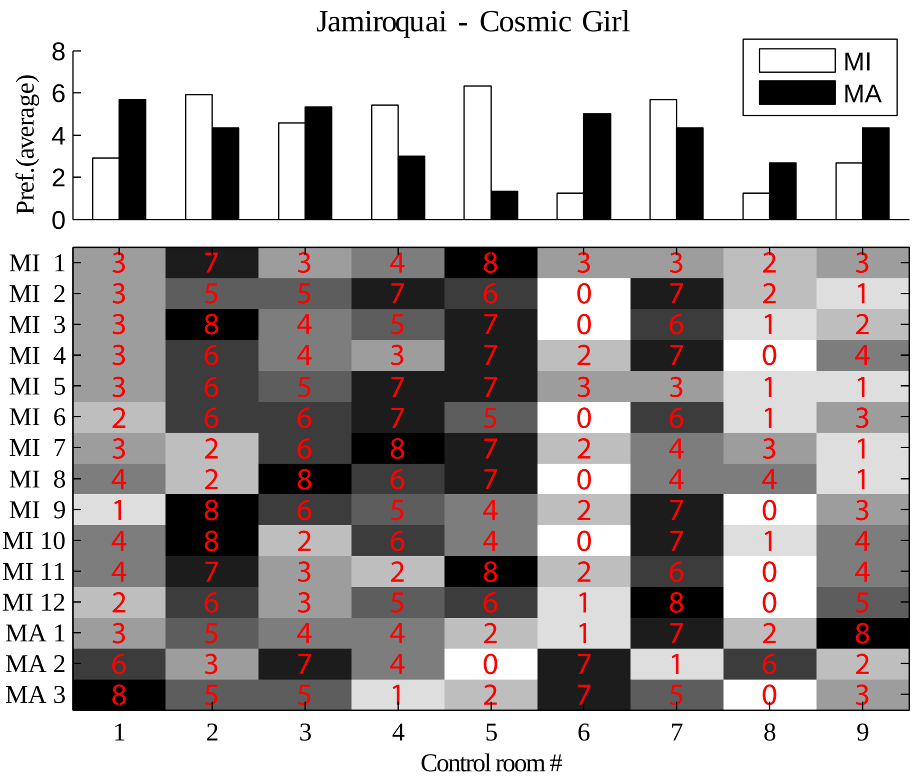


Figure 27: Preference rating chart on the basis of pair comparison test for the Sample 3 (Jamiroquai - Cosmic Girl). MI: mixing engineer, MA: mastering engineer.

MI 3:
"Firstly, I listened if I can localize the direction of the direct sound and perceive the distance to the loudspeakers. Based on my own experience, I have noticed that you can go easily wrong with balance adjustments if the control room has too long reverberation time. Also the stereo image was an important factor. The frequency balance came only after these two points."

MI 4:
"I think I made my decisions on the basis of vocals, especially whether vocals are in the middle and if they are near or far. In my listening, I emphasize a good and balanced stereo image. I think, I preferred mostly a stereo image which is mono (ed. narrow?). It (ed. The stereo-image) had to be somehow in a solid packet."

MI 5:
"I had two factors. How much detail do I hear and how much it tires me. Because that seems for me a bit of a balance. The more detail, the more it tires me. So it is a bit of a trade off for me between the clarity and how much it tires."

MI 6:
"The most important thing to me was the stability of the stereo image. I prefer that everything comes from one distinct plane. If the sounds seem to come from different

elevations, it makes the listening more difficult. The other thing was the distance to the sound sources. I think I preferred short reverberation times and loudspeakers that are nearer. I prefer that I perceive that the (ed. musical) material is close. After those things, I paid attention to the frequency balance and how low the sub-bass can reach."

MI 7:
"Stereo image was quite important. How the presence is perceived. And, for me, the low frequency reproduction is important, especially the clarity of low frequencies. Also the high frequencies were important. For me it is important to hear hi-hats clearly."

MI 8:
"I think the better the stereo image and the clearer you can separate different instruments, the better the room. Also the magnitude response should be flat, thus you can hear every frequency. I preferred shorter reverberation times."

MI 9:
"It depended quite much on the music sample. Stereo image was important and the presence of the singer. I prefer that the room ambience is not so dominant and that you about hear all frequencies."

MI 10:
"Directionality was the main thing. I did not give too much weight for the frequency balance. It is also important, but it was hard for me to rate it as these sound samples are not familiar to me."

MI 11:
"Firstly, the proximity of the mid-frequencies was important and also the clarity of the mid-frequencies. Second criteria was how reverberant the room is. Also the stereo image was important. With the Sample 2, I think all rooms succeeded quite well when it came to reverberance, but with Sample 3, I preferred more dry environments. It was also disturbing if there was happening too much in vertical axis."

MI 12:
"I think the frequency balance was the second priority. The most important for me was the reverberance, how the room colors the sound. I preferred drier rooms."

MA 1:
"The middle frequency balance is important for me first and foremost. I won't say that low and high frequencies are not important at all, but if I cannot make sense about the mid-frequencies, I think the control room has failed. I like some kind of proximity, but there must be a sense of surrounding spaciousness as well. I like this kind of picture of a stereo image where there is 2 dimensions. I do not need 3D effect in stereo image."

MA 2:
"Stereo balance and of course frequency balance. In a couple of rooms there appeared to be some phase issues. I mean that especially at upper mid frequencies phases sounded strange. I also prefer softness. There was quite harsh upper middle in many rooms that I did not prefer.

MA 3:
"A flat magnitude response was clearly the most important to me. I did not give too much weight for the amount of reverberation. I think that I preferred warmer frequency balance rather than too bright."

Question 2: How realistic you experienced the auralization? Could you think you were in a real control room when listening to the samples, and if not, what kind of deficiencies you found?

MI 1:
"I found it very cool how well this imitates the rooms. Very interesting experiment."

MI 2:
"I doubt if the shifting of localization in a certain rooms could be so significant in real situation. But overally it was quite realistic."

MI 3:
"I found differences between rooms surprisingly massive, even an overly massive. But I cannot say if that is because of the experiment situation or because of the unrealistic auralization. Because this situation allows you to change rooms without adaptation."

MI 4:
"I think that the most reverberant rooms sounded even too reverberant in the auralizations. Maybe this auralization also slightly exaggerates a certain deficiencies in rooms."

MI 5:
"I have not worked in these particular studios so I could not say if it sounded exactly like."

MI 6:
"Pretty good I think. Because I found both our control rooms. Very interesting experiment that should be done for every mixing engineer."

MI 7:
"I think this was realistic. It was fun when the loop ended, you could hear the actual

room reverberation and imagine the height and volume of the room."

MI 8:
–No comments–

MI 9:
"The most reverberant rooms I think sounded a bit like you listened through a reverb machine. I had some kind of 'machine' feeling at times. The driest rooms I found quite realistic though."

MI 10:
"It was quite realistic. Differences were transmitted very well. Our rooms sounded pretty much like in real situation."

MI 11:
"It was realistic. A couple of small things I noticed. For example that the center was not in center of the stereo image in a certain rooms. I doubt if it is like that in real rooms? Very cool experiment!"

MI 12:
"Very realistic! Can I listen a little bit more Within Temptation?"

MA 1:
"At least the rooms were extremely different sounding. I think that especially the width of the stereo image was successfully transmitted by this auralization"

MA 2:
"I found this extremely realistic. It was amazing that I even noticed the tone of my own loudspeakers although this auralization is reproduced by the Genelec's. From the beginning I was pretty sure which is my room."

MA 3:
"It was not so realistic. I think that rooms were too reverberant in this auralization. I also found the very high frequencies missing."

Question 3: Describe control rooms with your own words, what comes into your mind. You can listen rooms freely, but try to say a couple of words about each room.

CR#1:

Mixing engineers:
"Loudspeakers sounds to be quite far. Material (ed. music) sounds quite good in frontal area though. Very much reverberation"
"Very lively again. Still quite flat, but this reverberation is a little disturbing."

"This has a little bit night club feeling. I find this too reverberant."
"This is wide. No clear and accurate stereo image. I guess this room is quite large and has large loudspeakers as well. Frequency balance is nice. The room itself I find balanced, but I would like to have speakers more close."
"I feel this too wide. I cannot locate the singer properly. Low frequency reproduction I like though."
"I found this pleasant"
"The localization is not very good."
"This is a little bit problematic. Very wide. This is pleasant to listen but hard for mixing. Sides are very pronounced but the middle of the stereo image sounds a bit far."

Mastering engineers:
"I like this a lot."
"A good sense of space. Lots of good things. I like this."
"This sound extremely good. Quite wide stereo image but I like it. The punch of the bass drum is clearly audible."

CR#2:

Mixing engineers:
"Very good. Not much sound coming behind. However, there are some reflections to widen the stereo image in a good way. All genres sounds reasonable. Balanced sound. No colorations."
"I thinks this is quite good. Nice stereo image."
"Not bad. I think I could work in this room."
"I think stereo image is somehow tilted. Vocalist is shifted to the right. But this room is quite neutral."
"I think the stereo image is a bit shifted to the right. Clear high and low frequencies. I think this is good."
"The vocalist is in the middle as it should be."
"This sounds good. Quite balanced."

Mastering engineers:
"This is ok. Sounds a little bit distant and unclear in my taste."
"I think the 3 kHz area is a bit too crisp to me. It depends a lot about the genre if this works or not."

CR#3:

Mixing engineers:
"I hear some reflections coming back left all the time. Maybe some asymmetry with the room or something."

"This is a bit nasal. Some boost at 500 Hz or something. "
"Localization is a bit turned up in elevation."
"Vocalist is everywhere. You cannot point where the vocalist is. A little too much bass. I guess this is quite dry room but big speakers."
"This is quite pleasant. High frequencies could be more crispy. A little too wide stereo image for me."
"The localization is not so clear."
"A little too reverberant."

Mastering engineers:
"Quite good. Maybe a little bit treble boosted. Maybe also a little mid bass boomy."
"I found that there are some significant reflection affecting middle frequencies. Part of middle frequencies were missing. It felt strange to me."

CR#4:

Mixing engineers:
"Quite good spectral balance. But things spread a little bit from the stereo image. Maybe larger room."
"Pretty good."
"I feel like there are something coming from behind. But I do not think it would be a problem in mixing."
"A little boomy bass."
"This sounds a bit compressed in a way."
"Stereo image a bit turned to right. A bit too bassy."
"I do not like the high frequency reproduction of this."
"I think this is a mastering room. I like this."
"I felt the bottne is somehow loose from the other frequencies. But I like the sound though."

Mastering engineers:
"Slightly too much upper bass."
"I think this has a significant phase issues. Sounds like one driver is assembled to anti-phase."
"I did not like this somehow, do not know why though."

CR#5:

Mixing engineers:
"This is very pleasing. Extremely good sound. A little reverberation, everything localizes clearly to loudspeakers. Very good room."
"This is very clear."
"I find this a quite a good middle of different influences as in detail as in how much

it is tiring."
"Sound localizes pretty well to loudspeakers. Not much reflections. Maybe a little bit middle boosted spectral balance. Vocals could be still even more sharp."
"I like this stereo image."
"Quite pleasing tone."
"This is balanced"

Mastering engineers:
"I bet this has a small dip somewhere at low mids. Like 300 Hz or something. Also a little too much upper bass."
"I hear quite much middle boost here. Very harsh around 3-3.5 kHz. I could not work with this."

CR#6:

Mixing engineers:
"I think this is largest of all rooms. This has a significant reverberation. Frequency balance is pretty good though."
"This is very reverberant. I cannot control the hi-hat comp at all."
"I could not work with this. This sounds more like nightclub than a control room. I think, I would make mistakes in this room when mixing."
"This is quite pleasant to just listen to, but I think you could lose a bit detail."
"I feel I listen to more room than a mix. I cannot properly say where things are located in the mix."
"This loses the presence of the vocals. I think I could certainly not mix vocals here."
"I think this is pleasant to just listen to, but mixing in this would be hard."

Mastering engineers:
"This has a quite much reverberation. There are things I like in this though."
"This is quite ok. I would need a little more bottne (ed. low frequencies). I think at 100 Hz there is missing a certain punch."
"This has quite neutral frequency balance. Very much reverberation, but I could work here."

CR#7:

Mixing engineers:
"This is quite good when it comes to stereo image and reflections, but this has a slight upper bass boost."
"I feel the bottne (ed. low frequencies) is a little bit inaccurate."
"Quite balanced I think."
"Quite narrow stereo image."
"Stereo image is good."

"This feels quite balanced. This is a little bit narrower. Stereo image is convex compared to others. I mean middle of the stereo image is pronounced."
"This is really nice. I would mix here."

Mastering engineers:
"This is quite narrow. There is also a slight middle boost. Could be some reflection also."
"This is quite accurate, I could work here."

CR#8:

Mixing engineers:
"This has very much upper bass, say 150 - 200 Hz. Very much reverberation also. Sounds quite confusing. I cannot localize things. I think there is much early reflections in this room. On the other hand, this Sample 2 sounds very good in here."
"This is not accurate."
"It feels that left channel is odd compared to the right channel. Bottom end is not very accurate. Quite reverberant. Upper bass boosted."
"This is strange. Not accurate at all. I feel that loudspeakers are very far away."
"This is a bit muddy. A bit too much bottom end also."

Mastering engineers:
"Middle frequencies are ok. 100 Hz booms a bit."
"I like that robust bottom end, but I think there is some frequencies missing from the high end."
"There is some low frequency boom. And also the localization is a bit lost."

CR#9:

Mixing engineers:
"This is quite pleasing. The room a bit widens the stereo image. Frequency balance is a bit light. I would need more bottom end (ed. low frequencies)."
"Insufficient bottom end. I bit too reverberant also."
"I can hear the room all time. This does not localize to the loudspeakers very well."
"This is a bit too roomy. Stereo image not so clear."
"This is quite middle boosted. Stereo image is a bit messy."
"Insufficient bottom end for me."
"Bottom end is missing. I do not like this"
"Too much reverberation at middle frequencies maybe."
"I think every song sounds the same in this room. This sound very neutral room."
"I think the loudspeakers are quite small."

Mastering engineers:
"Many good things in this. Very simple and easy to internalize. Bottom end is attenuated though."
"Quite thin. I think all good bottom end and lower middle frequencies are missing."
"Weak bottom end but otherwise quite good."

6 Discussion

6.1 Connections between spatiotemporal visualizations and listening tests

Considering the preference of mixing engineers, Figs. 25, 26 and 27 show that CR#5 is preferred most with Sample 3 as well as with Sample 2. Closer inspection on the spatiotemporal visualization of CR#5 in Figs. 20 and 21 shows that early reflections are divided quite diffusively and no significant specular reflections exist in lateral plane. Also the 'spikes' of the direct sound are quite narrow in lateral plane. These could be the reasons for accurate stereo image, clarity and further to the high preference rating. Undoubtedly the reverberation time and flatness of the magnitude response has also impact to the high preference rating.

In Fig. 21 f), the table reflections of CR#5 are visible, as they are coming within 5 ms after the direct sound. Reflections have almost the same amplitude as the direct sound, and they are likely caused by the racks mounted to the both sides of the table. However, these reflections arrive quite early and from a narrow angle, and at least on the basis of the interviews, they were not reported to be disturbing. However, it must be kept in mind, that in the interview situation it could be hard to specify the reasons for disturbing effects.

Also CR#2, CR#3 and CR#7 were ranked high in the preference test by the mixing engineers. Figures 30 and 31 show that the amount of early reflections is quite small in CR#2 and CR#3. However, strong table reflection can be seen in CR#2, which surprisingly was not reported disturbing by mixing engineers. Although CR#7 has relatively high level of early energy coming from the frontal area, it has the shortest reverberation time, which is likely the reason for high preference rating. Based on the interview, CR#7 was reported to sound quite mono, which is due to slightly smaller angle between loudspeakers, but probably also due to the very early energy coming between the loudspeakers. This can be seen in Fig. 20. Also the shape of the cumulative response is more elliptic than round, thus early reflections are not diffuse, which can strengthen the sense of mono stereo image.

Figure 28 shows the lateral spatiotemporal visualization of the most and the least preferred control rooms among mixing engineers with Sample 3. It can be seen that the most prominent difference between most preferred control rooms (a and b) and least preferred control rooms (c and d) is the amount of early reflections compared to the direct sound. In Fig. 28 d), it can be seen also prominent asymmetry in early reflections. In addition, the spikes of the direct sound are more clearly distinguishable in a) and b) than in c) and d). Undoubtedly the significant ceiling reflection of the CR#6 has effect to the poor preference rating as well. The ceiling reflection can be seen in the Fig. 34 b). Examining the relationship between listening tests and spatiotemporal visualizations, conclusion can be made as follows. Generally it seems that when the early reflections exceed the level of -15 dB related to the direct sound, early reflections are unevenly distributed or significant amount of early energy is coming from the ceiling, the room is not preferred by mixing engineers.

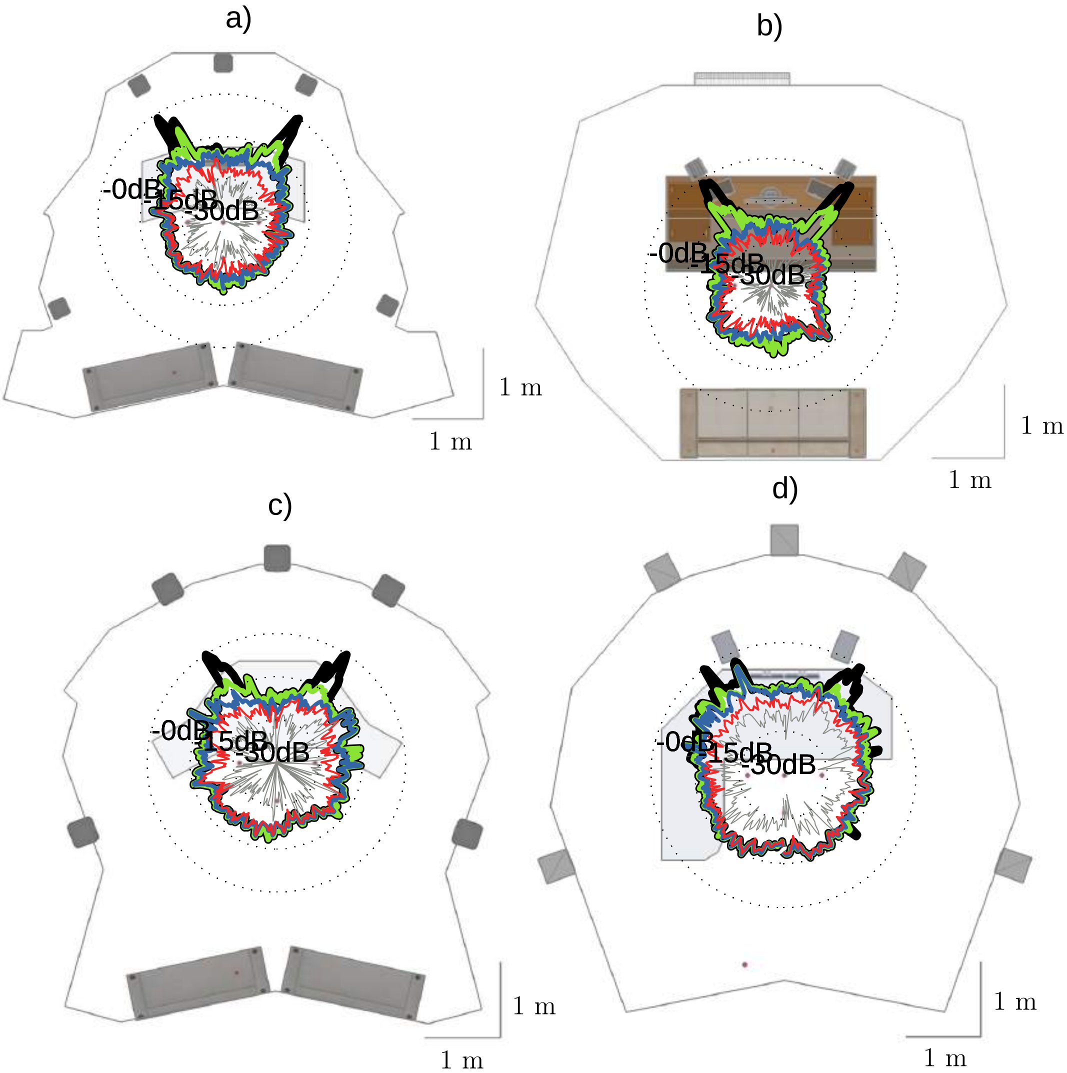


Figure 28: Cumulative polar response of two most preferred control rooms (a and b) and two least preferred control rooms (c and d) among mixing engineers with the sample 3. a) CR#5, b) CR#2, c) CR#6 and d) CR#8.

6.2 Credibility of the analysis methods

Spatiotemporal visualizations give accurate and useful information about the early reflected energy in control rooms. Despite of the possible localization errors discussed in [54], certain logical connections between perceptual effects and the visualizations can be observed. The width and the accuracy of the stereo image can be predicted from the visualizations as discussed in Chapter 6.1. In addition, spatiotemporal plots show intuitively the ratio of direct sound and early reflections and also the shape how the cumulative early energy is divided within the room. These aspects are very important when designing and studying the control room acoustics.

In the interview, subjects were asked whether they experienced the auralization realistic or not. It can be concluded from the citations presented in Chapter 5.2.2

that majority of the sound engineers found auralizations realistic, and that the auralizations are preserving the key features of acoustics in control rooms. Several subjects noted that the most reverberant control rooms sounded too reverberant. This is probably due to the fact that subjects adapted to the anechoic chamber, and auralizations were perceived unnatural when there is no possibility to adapt to the room itself before listening to each sample. Another reason for the perceived excessive reverberation could be the one noted in [54], that the SDM seems to increase the late reverberation in some cases. This is because SDM hypothesis assumes that there is only one reflection present in each analysis window. This assumption is not always true in small spaces since the echo density increases very quickly with respect to time.

Another issue that a couple of subjects reported in the interview, was that in auralizations the center image is not in the physical center between the ±30 degrees lateral loudspeakers. This was noticed as the vocalist was not in the middle where it should be. One possible explanation for this could be that the left and the right loudspeakers actually reproduced different magnitude response to the listening position in certain rooms. Figure 43 shows that in many rooms there are SPL differences between left and right loudspeaker at certain frequencies. Another reason can be the possible minor errors in the microphone placement during the measurements. Thus, it is possible that the microphone array has not been exactly in the center between the loudspeakers in every measurement, because the positioning was done by hand. Yet another reason can be the fact that auralizations where implemented with one single measurement point. In reality, music is not listened only in one point, but slight movement of head occur. This movement possibly help to solve the cues of localization and the virtual center is perceived differently than in auralizations.

Undoubtedly, the visual aspects play a key role in the localization as well. In the anechoic chamber there is loudspeakers all over the room and in control room there are only two speakers and engineer knows exactly from which direction the sound is coming. A well-known fact is, that the sound perception mechanism in brains combines the auditory and visual cue when constructing the directional sensation of the sound [64]. Thus, in an anechoic chamber, where sound sources are not unambiguous, existing image shifts are more easily perceived than in real rooms where visual cues shape the localization perception as the two stereo loudspeakers are usually clearly visible.

Ten of the 15 sound engineers who participated in the listening test, had their own control room included in the study. During the interview, sound engineers were asked if they can recognize their own room from the nine alternatives. Almost every sound engineer recognized their room, and most of them were quite sure about it. A couple of engineers could not point the exact room, because of the sound samples were not familiar to them. However, this result confirms the fact that the auralization was succesful.

As it can be noticed from the interview results, there are a few contradictory comment within the same control room. Thus, it can be deduced that whether some of the subjects could not give a consistent answers or different subjects pursue different attributes way differently. To eliminate the effect of these kind of inconsistencies, the listening tests would have needed more participants, or the listening tests should be conducted in a form of Individual Vocabulary Profiling (IVP) test [35], in which subjects form their own attributes and listen to each attribute at a time. However, this kind of extensive listening test was not possible within the scope of this thesis.

6.3 Role of the loudspeakers

In this work, the listening conditions of control rooms were decided to explore as they currently exist. This means that the own monitor loudspeakers of each control room were used as a sound source in the measurements. This must be taken into account when interpreting the results of the listening tests. Firstly, sound engineers have a different preferences about which loudspeaker best fulfill their needs. Others prefer extremely flat and neutral loudspeakers and others need the loudspeakers to have some imperfections or tone to get the job done. For example in CR#9, the magnitude response is deliberately adjusted so that the low frequencies are attenuated, as can be seen in the Figs. 42 and 37 d). This was reported by almost every subject, and they did not prefer the room for that reason. This does not mean that the room itself has any problems or defects.

The placement and orientation of the loudspeakers were also variables which turned out to divide the sound engineers. Others preferred a more wide stereo placement of the loudspeakers and others narrower. On the other hand, majority of the sound engineers preferred near field setup, but in contrast, a few reported to like far field monitoring. These facts must be kept in mind when making conclusions about the rooms and searching connections between preference ratings and the acoustic measures, such as spatiotemporal visualizations.

6.4 Practical implications of the presented work

Significant note related to the practical utility of this work is that many studio owners performed, or at least planned, renovations or re-arrangements in their control room based on the results of this work. For example, the strong ceiling reflections that the spatiotemporal visualizations revealed in CR#6, turned out to be due to an installation error, and actions are taken to repair the problem. In addition, in CR#9, the listening position was moved slightly back on the basis of the measurement results of this work and the level of the low frequencies was increased.

6.5 Review of the specifications for an ideal control room

The results of this work show that mixing engineers prefer rooms with T_{60} between 0.17 and 0.26 s, which is slightly less than presented in Chapter 3.5. For mastering engineers, preferred T_{60} seems to be between 0.3 and 0.4 s, which is significantly more than with mixing engineers.

Regarding the room modes, the results of the listening tests support the specification that a room is successful if room modes are properly distributed and resonances damped. Figure 36 e) illustrates that there are several ringing modes in

CR#8, and that is the most probable reason for the 'booming' that many sound engineers reported in interview.

The specification of preventing reflections that have a level higher than -15 dB with respect to direct sound during the first 20 ms, is a little problematic. There does not seem to be much correlation between that specification and the preference ratings or interviews. One alternative for reviewing the magnitude of reflections would be to examine the energy during the first 20 ms. On the basis of the subjective listening tests conducted by Toole *et al.* [56, p. 90-91], the multiple low-level reflections were perceived equally loud as a single reflection. This advocates the examination of the cumulative energy instead of the amplitude of individual reflections.

On average, the magnitude response was not experienced as the most important factor on the basis of the listening tests. There were also different views among sound engineers about the optimal frequency balance. However, in the interviews, the majority of the mixing engineers mentioned flat magnitude response as a positive attribute.

Based on the results of this work, it seems that mastering and mixing needs different rooms. It can be also deduced that mastering engineers have different personal needs when it comes to the control room acoustics. This was pointed out also by Augspurger in [3]. Despite of the personal differences in control room preferences, reviewed general specifications of the optimal control room are presented as follows:

1. Flat magnitude response at least in the listening position (from 20 Hz to 20 kHz $\pm$ 4 dB). Equal magnitude response in the listening position for all loudspeakers.

2. Frequency balanced reverberation time of 0.20 s for mixing engineers and 0.35 s for mastering engineers.

3. Proper distribution of room modes to produce accurate low frequency reproduction.

4. Initial time delay gap of 20 ms, that is, after the direct sound there should be no reflections over -15 dB during the first 20 ms.

5. Symmetric early reflections to prevent image shifting in stereo.

6.6 Future work

Future work on this topic would include more extensive listening tests with more participants. Listening tests could be done with IVP method to better clarify the attributes that sound engineers are listening to, and to get more exact information how rooms differ between these attributes. Secondly, control room acoustics could be studied with virtual control room simulator, in which sound engineers can adjust the parameters, such as, reverberance, shape of the reverberation, amount of early lateral reflections and frequency balance, to best correspond their needs. This experiment would be possible using a real time convolutions and a multi channel listening setup in an anechoic chamber.

It would be also interesting to investigate the effects of pure room by measuring all control rooms using the same loudspeakers as a sound source. This would eliminate the effect of the loudspeaker as a variable, and enable to study the effects of the room.

Also the effects of different acoustic treatment would be useful to explore. This could be done by conducting a case study while building a control room. Spatial impulse response measurements could be made in every step of the building process. This would enable the auralization of the sound fields at every revision of the control room and to compare the effects of each piece of an acoustical treatment by listening.

7 Conclusions

The scope of this thesis was divided into two main parts. Firstly, this thesis presented the use of spatiotemporal visualizations to study the early reflections of studio control rooms. Spatial impulse responses of 13 rooms were measured and acoustic reflections tracked with SDM. Spatiotemporal visualizations of each control room-loudspeaker combination were further drawn and analyzed. In addition, the Matlab tool for more exact tracking of early acoustic reflections was implemented. Implemented visualizations turned out to give accurate and intuitive information about early sound fields of studio control rooms. Significant acoustic defects were found with implemented visualization tools, that would have been hard to notice with traditional objective measures.

Secondly, this thesis studied the preference of studio control rooms among professional sound engineers. To enable A/B comparison between different control room-loudspeaker combinations, rooms were auralized on the basis of SDM analysis. In the listening tests, auralizations were played back with 30 channel loudspeaker system in an anechoic chamber. Preference test was conducted in a form of pair comparison, where subjects listened the auralizations, and chose the room in which they would prefer to work. After the preference tests, subjects were interviewed to reveal their arguments behind the preference decisions. In an interview, each control room was also listened individually, and mixing engineers asked to describe the acoustics of each room with their own words.

Finally, the results of the listening tests and their connections to spatiotemporal visualizations were discussed. The results of the preference tests clearly showed that mixing engineers prefer quite dry rooms (T_{60} of 0.15 - 0.20 s) and interviews confirm that the stereo image and the amount of room reverberation are the most important factors for them. In contrast, mastering engineers seemed to prefer more lively rooms (T_{60} of 0.30 - 0.40 s) and the frequency balance was the most important factor for them. It was also noticed that the preference rating varied between different music samples, especially among mixing engineers.

Concerning the specifications for an ideal control room, it seems that the earlier studies are mostly in line with the results obtained in this work. However, there are several factors to be emphasized. Firstly, the accuracy of the stereo image was experienced very important by almost every mixing engineer in this study. Secondly, the optimal reverberation time seemed to be yet shorter than earlier studies showed. Finally, on the basis of the listening tests, it seemed that different music genres need different kind of treatment in control room.

Based on the obtained results and the input from the professional sound engineers, it can be said that both visualizations and auralizations were successful. Overall, the work in this thesis achieved its objectives and new information was acquired about critical listening environments as well as about the preference in control room acoustics among sound engineers.

References

[1] ISO 3382, Acoustics—Measurement of the reverberation time of rooms with reference to other acoustic parameters. International Standards Organisation, 1997.

[2] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small room acoustics. *The Journal of the Acoustical Society of America*, 65(4), 1979.

[3] G. L. Augspurger. Control room design: The monitoring environment. *Acoustics Today*, 9(2):15–20, 2013.

[4] M. Barron. The subjective effects of first reflections in concert – the need for lateral reflections. *Journal of Sound and Vibration*, 15(4):475 – 494, 1971.

[5] A. Baskind and J. D. Polack. Sound power radiated by sources in diffuse field. In *Audio Engineering Society Convention 108*, Feb 2000. Paper number 5146.

[6] J. W. Bayless. Innovations in studio design and construction in the capitol tower recording studios. *J. Audio Eng. Soc*, 5(2):71–76, 1957.

[7] L. L. Beranek. The notebooks of Wallace C. Sabine. *The Journal of the Acoustical Society of America*, 61(3):629–639, 1977.

[8] L. L. Beranek. *Concert halls and opera houses: music, acoustics, and architecture.* Springer, Berlin, Germany, 2004.

[9] Mr. Bonzai. George massenburg builds a blackbird room. *Digizine, Digidesign.*

[10] J. Borish and J. B. Angell. An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. Audio Eng. Soc*, 31(7/8):478–488, 1983.

[11] L. J. Brannmark, A. Bahne, and A. Ahlén. Compensation of loudspeaker–room responses in a robust MIMO control framework. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(6):1201–1216, 2013.

[12] J. D. Bunton and R. H. Small. Cumulative spectra, tone bursts, and apodization. *J. Audio Eng. Soc*, 30(6):386–395, 1982.

[13] T. J. Cox and P. D'Antonio. *Acoustic absorbers and diffusers: theory, design and application.* Routledge Taylor & Francis, London, UK, 2009.

[14] P. D'Antonio and J. H. Konnert. The RFZ/RPG approach to control room monitoring. In *Audio Engineering Society Convention 76*, 10 1984. Paper number 2157.

[15] Z. Davidovic and B. Petrovic. Acoustical design of control room for stereo and multichannel production and reproduction - a novel approach. In *Audio Engineering Society Convention 129*, 11 2010. Paper number 8295.

[16] D. Davis. The role of the initial time delay gap in the acoustic design of control rooms for recording and reinforcement systems. In *Audio Engineering Society Convention 64*, Nov 1979. Paper number 1547.

[17] D. Davis and C. Davis. The LEDE- concept for the control of acoustic and psychoacoustic parameters in recording control rooms. *J. Audio Eng. Soc*, 28(9):585–595, 1980.

[18] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*, Feb 2000. Paper number 5093.

[19] B. Fazenda and W. J. Davies. The views of recording studio control room users. *Proceedings of the Institute of Acoustics*, 23(8):1–8, 2002.

[20] B. Fazenda, K. R. Holland, P. R. Newell, and S. V. Castro. Modulation transfer function as a measure of room low frequency performance. *Proceedings of the Institute of Acoustics*, 28(8):187–194, 2007.

[21] B. M. Fazenda and W. J. Davies. The views of recording studio control room users. In *Institute of Acoustics*, pages 1–8, 2001.

[22] B. M. Fazenda, K. R. Holland, and P. R. Newell. Measurement of sound quality for critical listening. In *Proceedings of 12th International Congress on Sound and Vibration*, 2005.

[23] Y. Fukushima, H. Suzuki, and A. Omoto. Visualization of reflected sound in enclosed space by sound intensity measurement. *Acoustical science and technology*, 27(3):187–189, 2006.

[24] D. Hearn and M. Pauline Baker. *Computer graphics.* Prentice-Hall, Inc., New Jersey, USA, 1997.

[25] E. J. Heller. *Why You Hear What You Hear: An Experiential Approach to Sound, Music, and Psychoacoustics.* Princeton University Press, New Jersey, USA, 2012.

[26] H. Imamura, A. Marui, T. Kamekawa, and M. Nakahara. Influence of first reflections in listening room on subjective listener impression of reproduced sound. In *Audio Engineering Society Convention 134*, 5 2013. eBrief number 82.

[27] ITU. ITU-R BS.775-2, Multichannel Stereophonic Sound System with and without Accompanying Picture. ITU Radiocommunication Assembly, 2006.

[28] B. Ivancevic, H. Domitrovic, and S. Fajt. Acoustical properties of control rooms. In *Audio Engineering Society Convention 102*, 3 1997. Paper number 4424.

[29] R. Izhaki. *Mixing Audio – Concepts, Practices and Tools.* Focal Press, Oxford, UK, 2008.

[30] M. Karjalainen. *Kommunikaatioakustiikka.* Helsinki University of Technology, Espoo, Finland, 1999.

[31] R. A. Katz. *Mastering Audio: The Art and the Science.* Focal Press, Newton, MA, USA, 2003.

[32] R. King, B. Leonard, and G. Sikora. The practical effects of lateral energy in critical listening environments. In *Audio Engineering Society Convention 131*, 10 2011. Paper number 8565.

[33] H. Kuttruff. In *Room Acoustics - Fifth Edition.* Spon Press, London, UK, 2009.

[34] B. Leonard, R. King, and G. Sikora. The effect of acoustic environment on reverberation level preference. In *Audio Engineering Society Convention 133*, Oct 2012. Paper number 8742.

[35] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo. Concert hall acoustics assessment with individually elicited attributes. *Journal of the Acoustical Society of America*, 130(2):835–849, August 2011.

[36] S. K. Mitra. *Digital Signal Processing: A Computer-Based Approach.* McGraw-Hill Higher Education, Columbus, USA, 2nd edition, 2000.

[37] A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki. Modal equalization of loudspeaker - room responses at low frequencies. *J. Audio Eng. Soc*, 51(5):324–343, 2003.

[38] M. Nakahara and A. Omoto. Room acoustic design for small multichannel studios. In *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, 6 2003.

[39] P. Newell. *Recording Studio Design, Third Edition.* Focal Press, Waltham, USA, 2012.

[40] S. H. Nielsen. Auditory distance perception in different rooms. *J. Audio Eng. Soc*, 41(10):755–770, 1993.

[41] S. E. Olive, P. L. Schuck, S. L. Sally, and M. Bonneville. The variability of loudspeaker sound quality among four domestic-sized rooms. In *Audio Engineering Society Convention 99*, Oct 1995.

[42] S. E. Olive and F. E. Toole. The detection of reflections in typical rooms. In *Audio Engineering Society Convention 85*, Nov 1988.

[43] A. Omoto and H. Uchida. Evaluation method of artificial acoustical environment: Visualization of sound intensity. *Journal of physiological anthropology and applied human science*, 23(6):249–253, 2004.

[44] J. Pätynen, S. Tervo, and T. Lokki. Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses. *J. Acoustical Society of America*, 133(2):842–857, Feb. 2013.

[45] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc*, 45(6):456–466, 1997.

[46] M. T. Putnam. Recording studio and control room facilities of advance design. *J. Audio Eng. Soc*, 8(2):111–119, 1960.

[47] M. T. Putnam. A thirty-five year history and evolution of the recording studio. In *Audio Engineering Society Convention 66*, May 1980. Paper number 1661.

[48] M. R Schroeder. Diffuse sound reflection by maximum- length sequences. *The Journal of the Acoustical Society of America*, 57:149, 1975.

[49] M. R. Schroeder. Modulation transfer functions: Definition and measurement. *Acta Acustica united with Acustica*, 49(3):179–182, 1981.

[50] M. R. Schroeder. Statistical parameters of the frequency response curves of large rooms. *J. Audio Eng. Soc*, 35(5):299–306, 1987.

[51] M. R. Schroeder, T. D. Rossing, F. Dunn, W. M. Hartmann, D. M. Campbell, and N. H. Fletcher. *Springer Handbook of Acoustics*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[52] Audio Engineering Society. Multichannel surround sound systems and operations. Technical report, 2001.

[53] G. B. Stan, J. J. Embrechts, and D. Archambeau. Comparison of different impulse response measurement techniques. *J. Audio Eng. Soc*, 50(4):249–262, 2002.

[54] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki. Spatial decomposition method for room impulse responses. *J. Audio Eng. Soc*, 61(1/2):17–28, 2013.

[55] F. E. Toole. Loudspeakers and rooms for sound reproduction, a scientific review. *J. Audio Eng. Soc*, 54(6):451–476, 2006.

[56] F. E. Toole. *Sound Reproduction: Loudspeakers and Rooms*. Focal Press. Elsevier, Amsterdam, Netherlands, 2008.

[57] F. E. Toole. Music, rooms and listeners science in the creation and delivery of audio art. *Acoustics Today*, 9(2), 2013.

[58] J. Voetmann. 50 years of sound control room design. In *Audio Engineering Society Convention 122*, 5 2007. Paper number 7140.

[59] E. J. Völker. Acoustics in control rooms-that recurring, burdensome subject. In *Audio Engineering Society Convention 105*, Sep 1998. Paper number 4832.

[60] R. Walker. A new approach to the design of control room acoustics for stereophony. In *Audio Engineering Society Convention 94*, 3 1993. Paper number 3543.

[61] R. Walker. Early reflections in studio control rooms: The results from the first controlled image design installations. In *Audio Engineering Society Convention 96*, 2 1994. Paper number 3853.

[62] R. Walker. Equalization of room acoustics and adaptive systems in the equalization of small room acoustics. In *Audio Engineering Society Conference: 15th International Conference: Audio, Acoustics & Small Spaces*, Oct 1998.

[63] R. Walker. Room acoustics for multichannel listening: Early reflection control. In *Illusions in sound - AES 22nd UK Conference*, pages 8–1–8–16, 2007.

[64] H. A. Witkin, S. Wapner, and T. Leventhal. Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology*, 43(1):58, 1952.

[65] H. D. Young and R. A. Freedman. *University Physics - Tenth Edition.* Addison Wesley Longman Inc., Boston, USA, 2000.

[66] D. Zhang and X. Wu. On cross correlation based-discrete time delay estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing.*, volume 4, pages iv/981–iv/984 Vol. 4, 2005.

A Measurement results (R1)

Each figure in appendix A contains following subfigures: Spatiotemporal visualization in (a) lateral plane and (b) median plane. Different colors correspond different time windows as presented in Fig. 13. To reveal the dimensions of control rooms, 1 m x 1 m scale is included in bottom right corner of every spatiotemporal visualization, as seen in both subfigures (a) and (b). (c) illustrates ETC summed from left and right stereo channels. (d) is magnitude response, where red color corresponds left channel and blue is right channel. (e) is CSD plot summed from left and right stereo channels. In CSD, energy is plotted as a function of both time and frequency. Note that in CSD, the frequency axis is limited to cover only low frequencies (from 20 Hz to 500 Hz).

A.1 CR#1

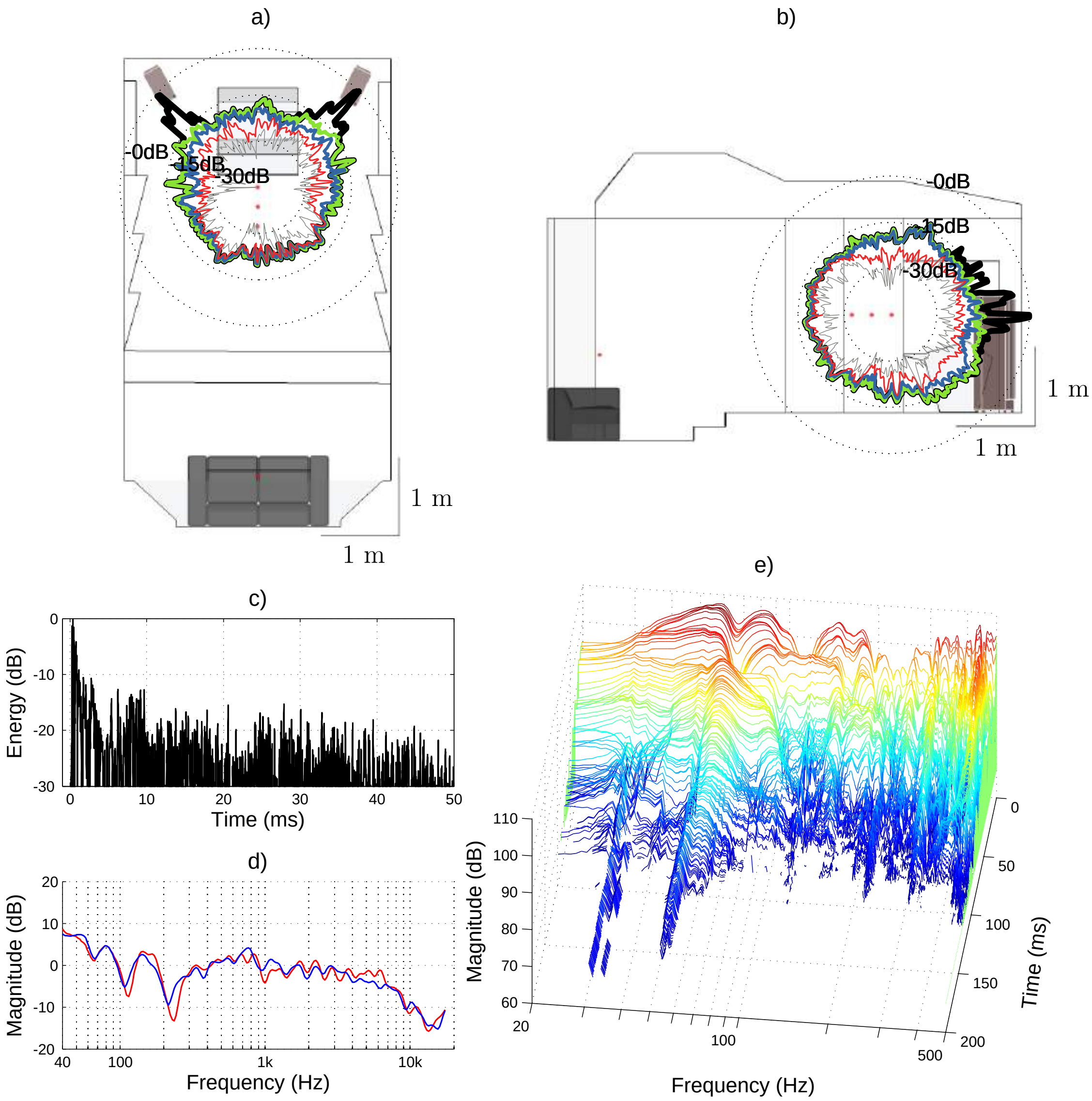


Figure 29: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#1.

A.2 CR#2

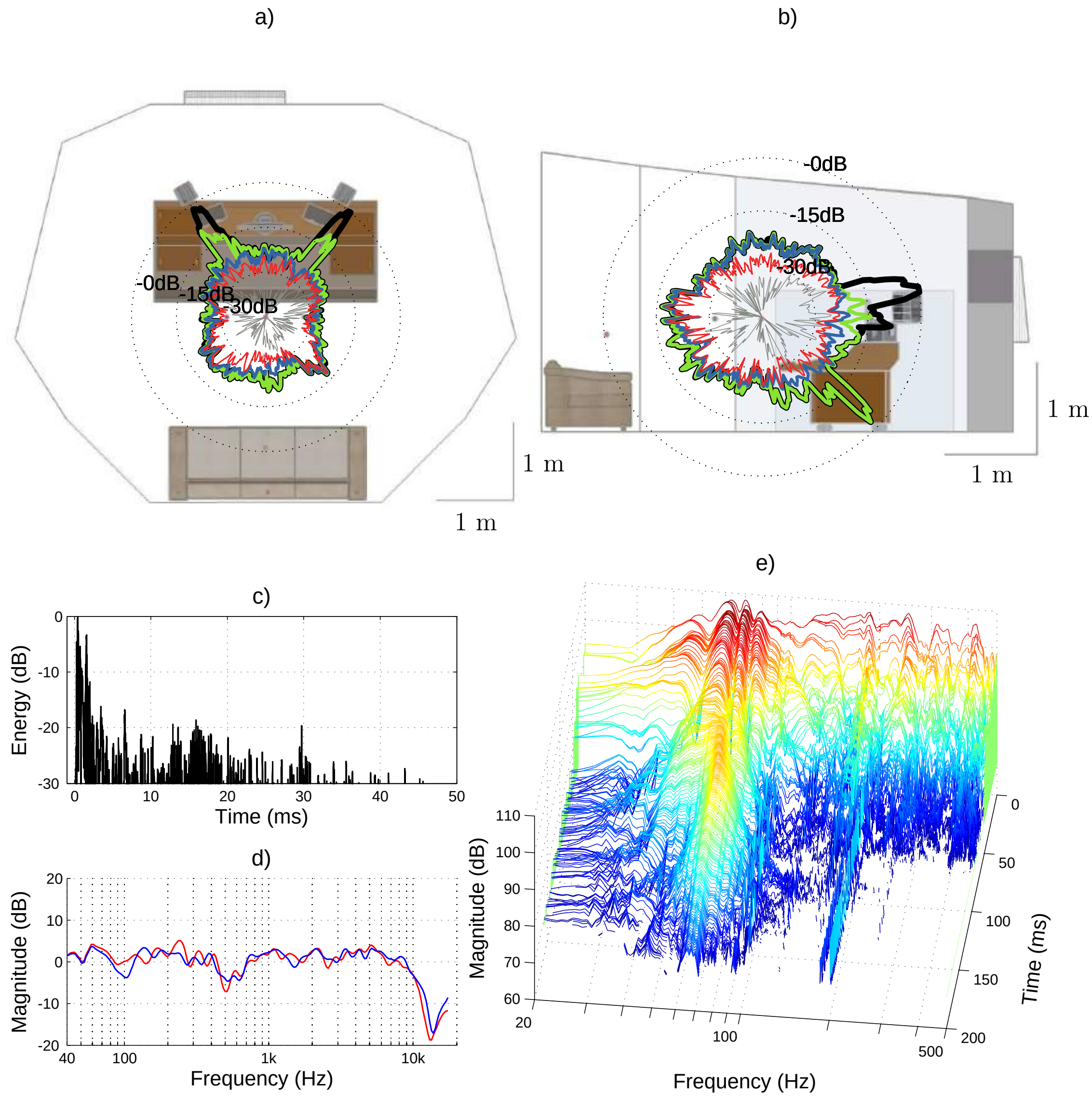


Figure 30: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#2. Note that CR#2 utilizes the same room as CR#3, but the difference is that CR#2 uses near field monitors, and CR#3 flush mounted monitor loudspeakers.

A.3 CR#3

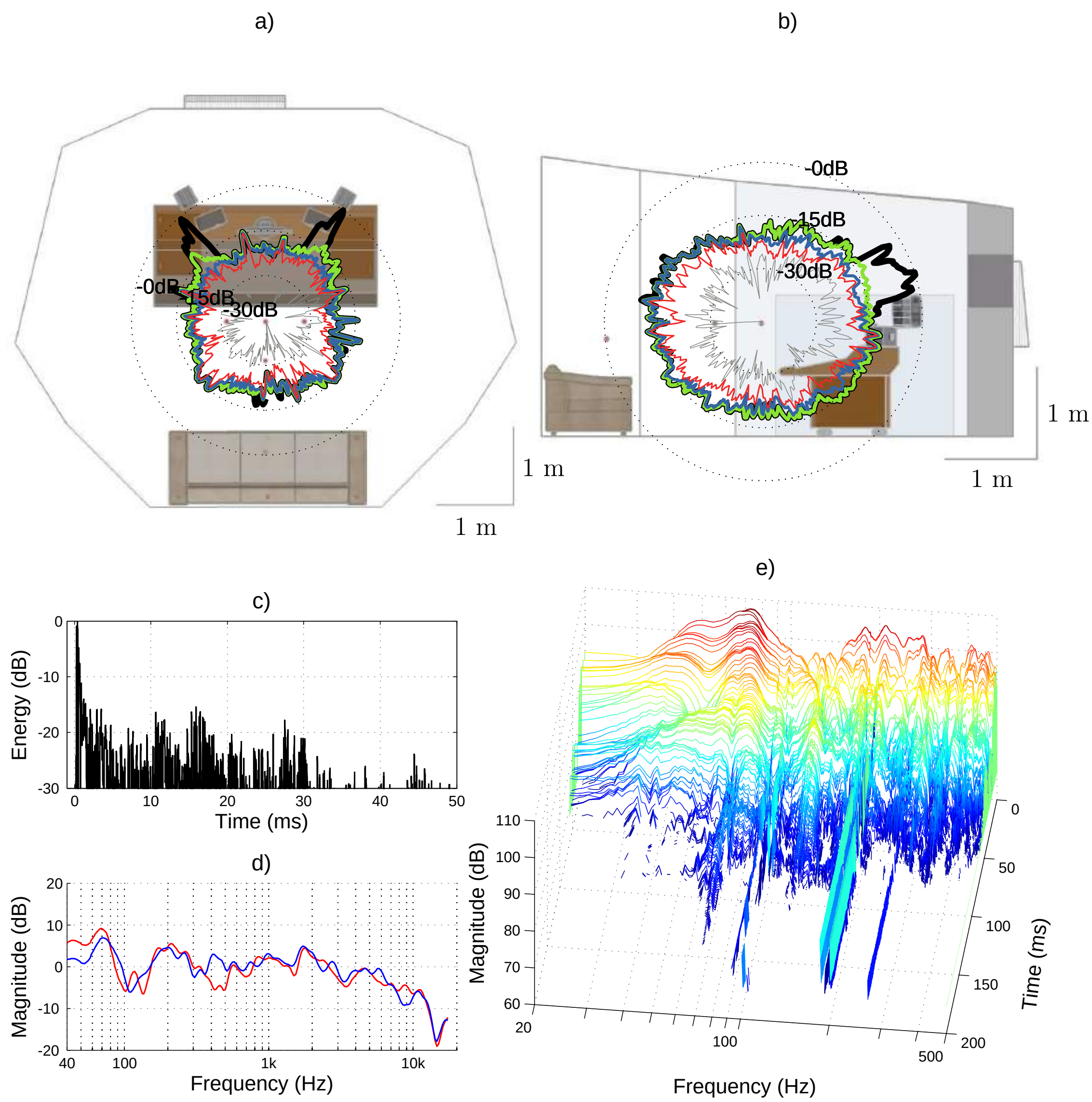


Figure 31: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#3. Note that CR#2 utilizes the same room as CR#3, but the difference is that CR#2 uses near field monitors, and CR#3 flush mounted monitor loudspeakers.

A.4 CR#4

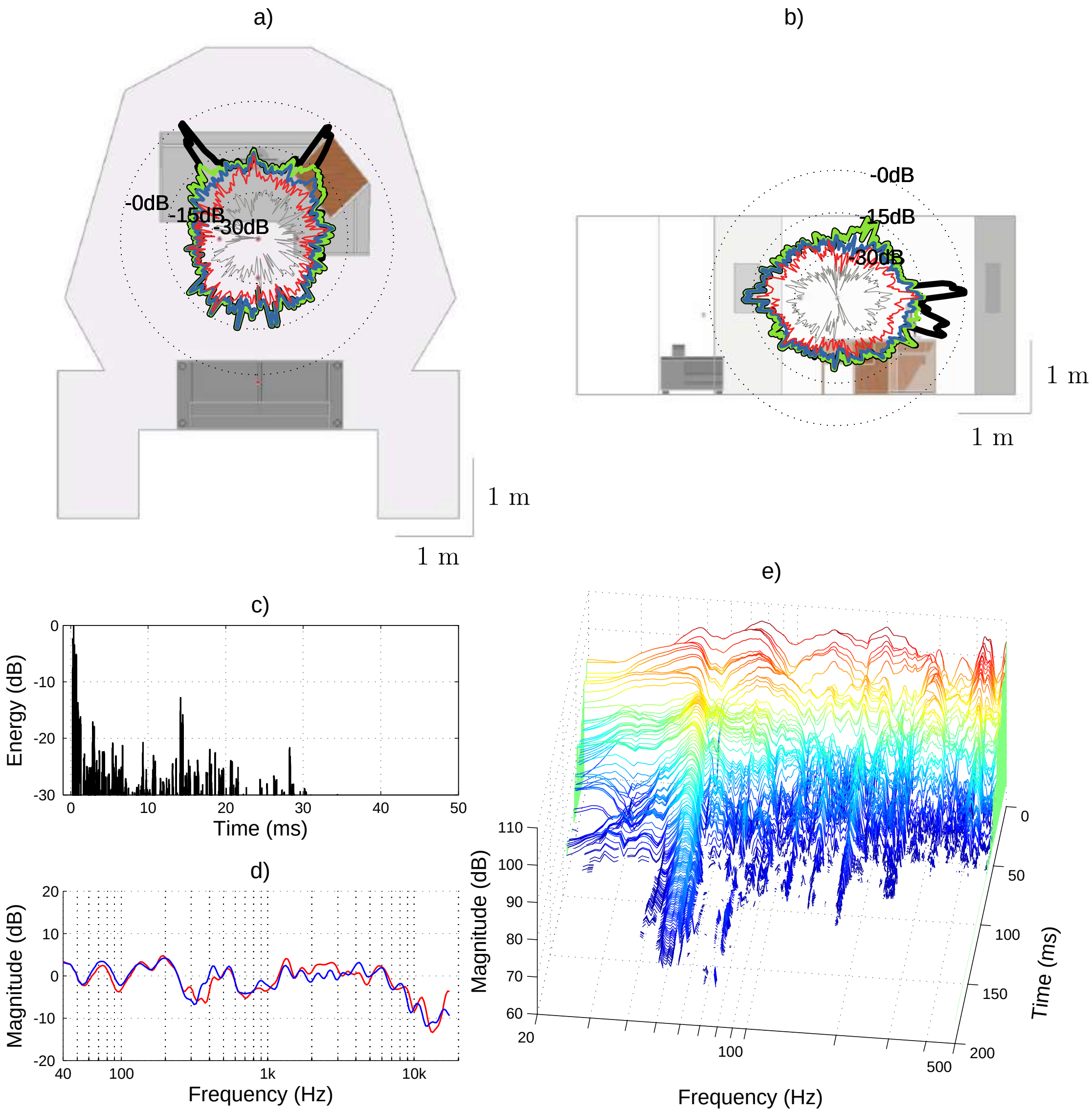


Figure 32: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#4.

A.5 CR#5

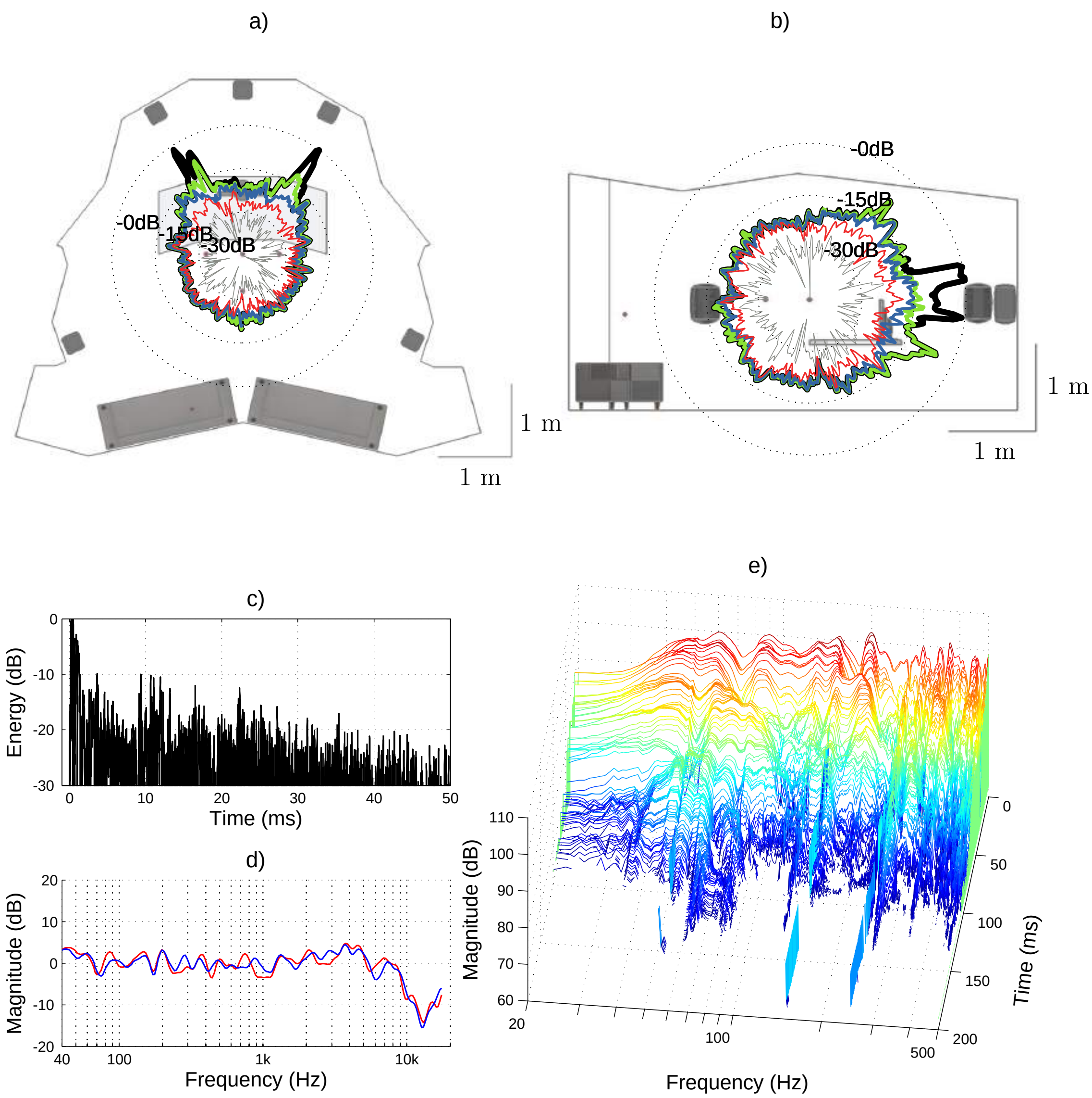


Figure 33: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#5.

A.6 CR#6

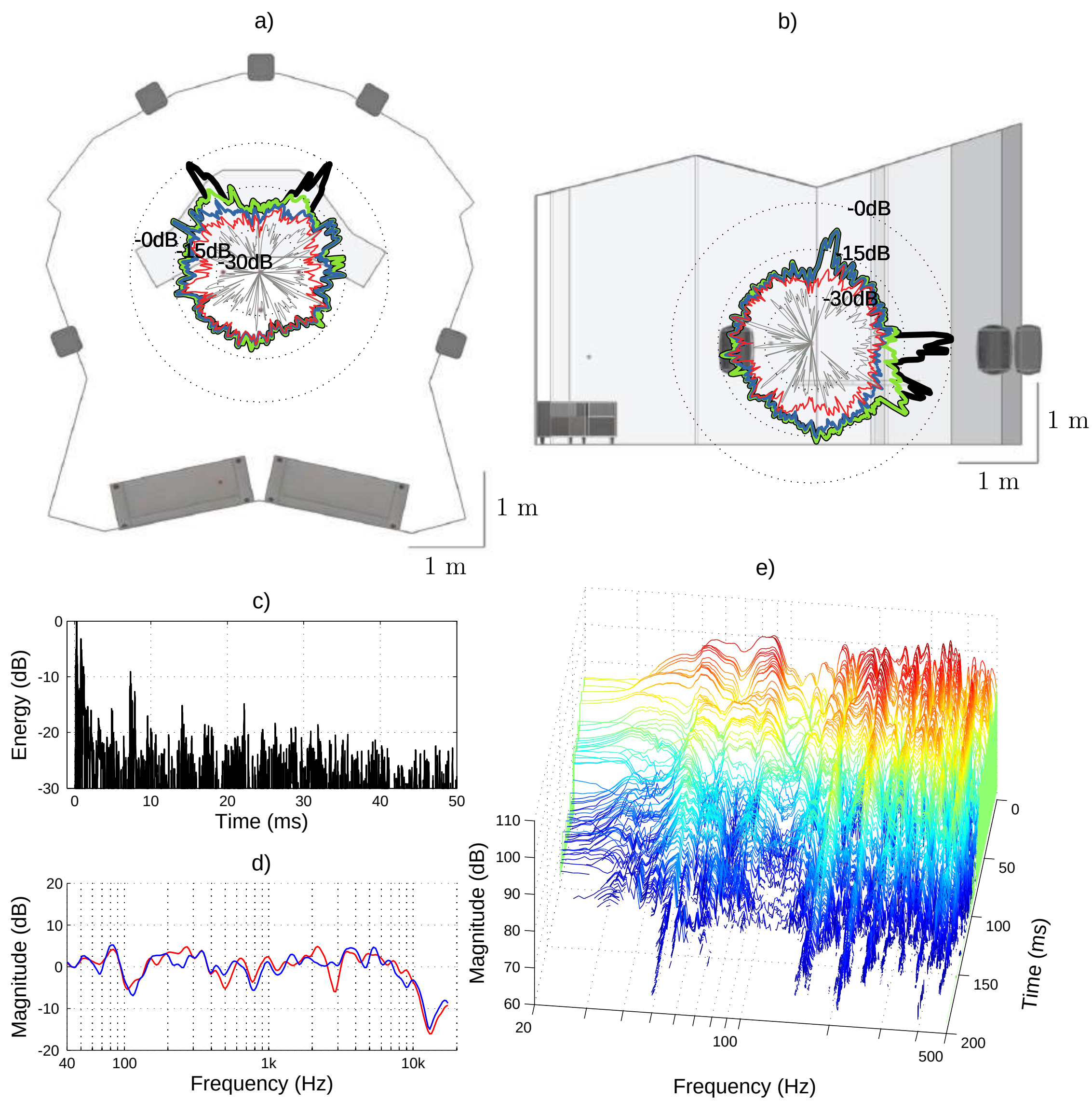


Figure 34: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#6.

A.7 CR#7

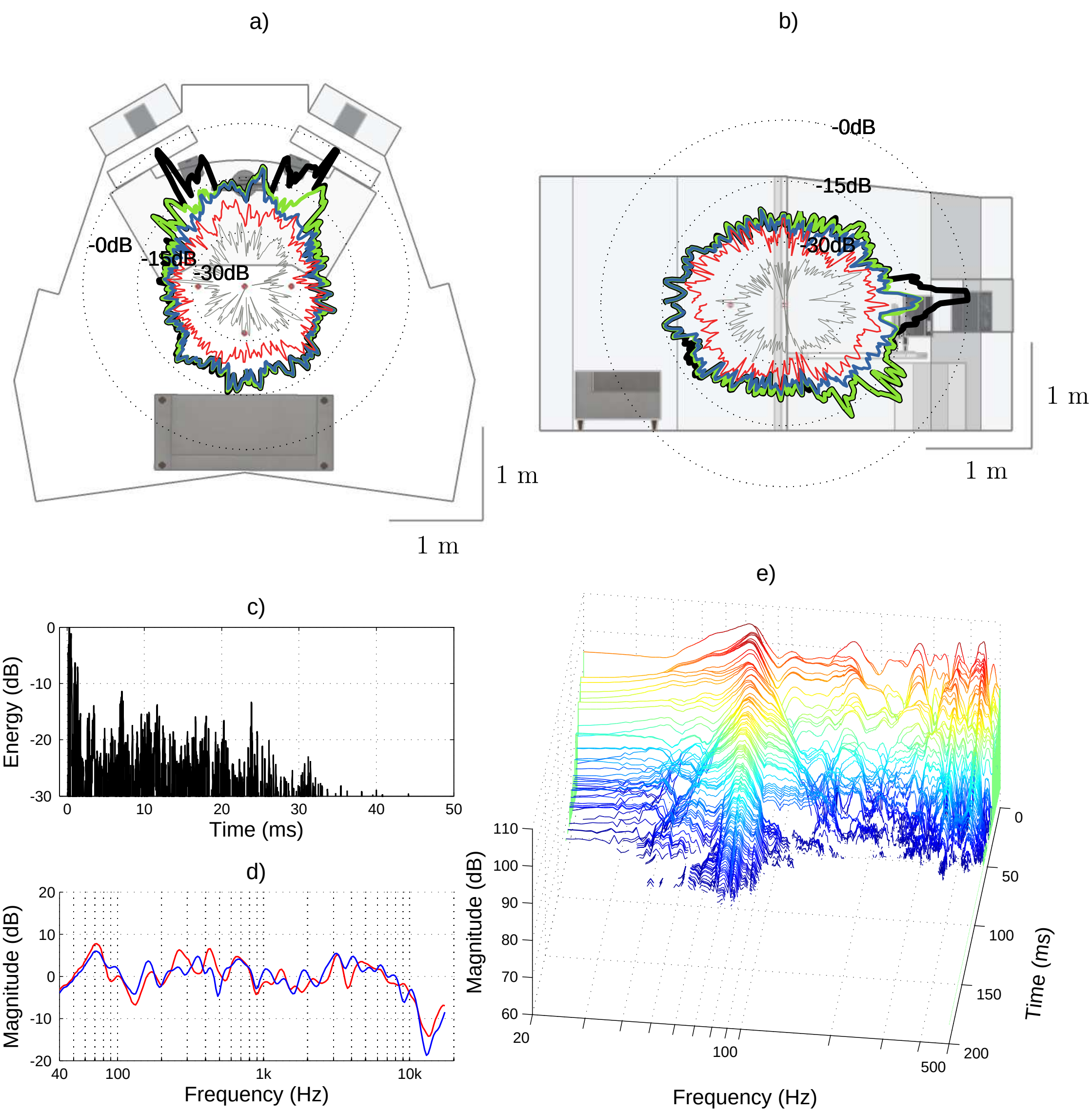


Figure 35: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#7.

A.8 CR#8

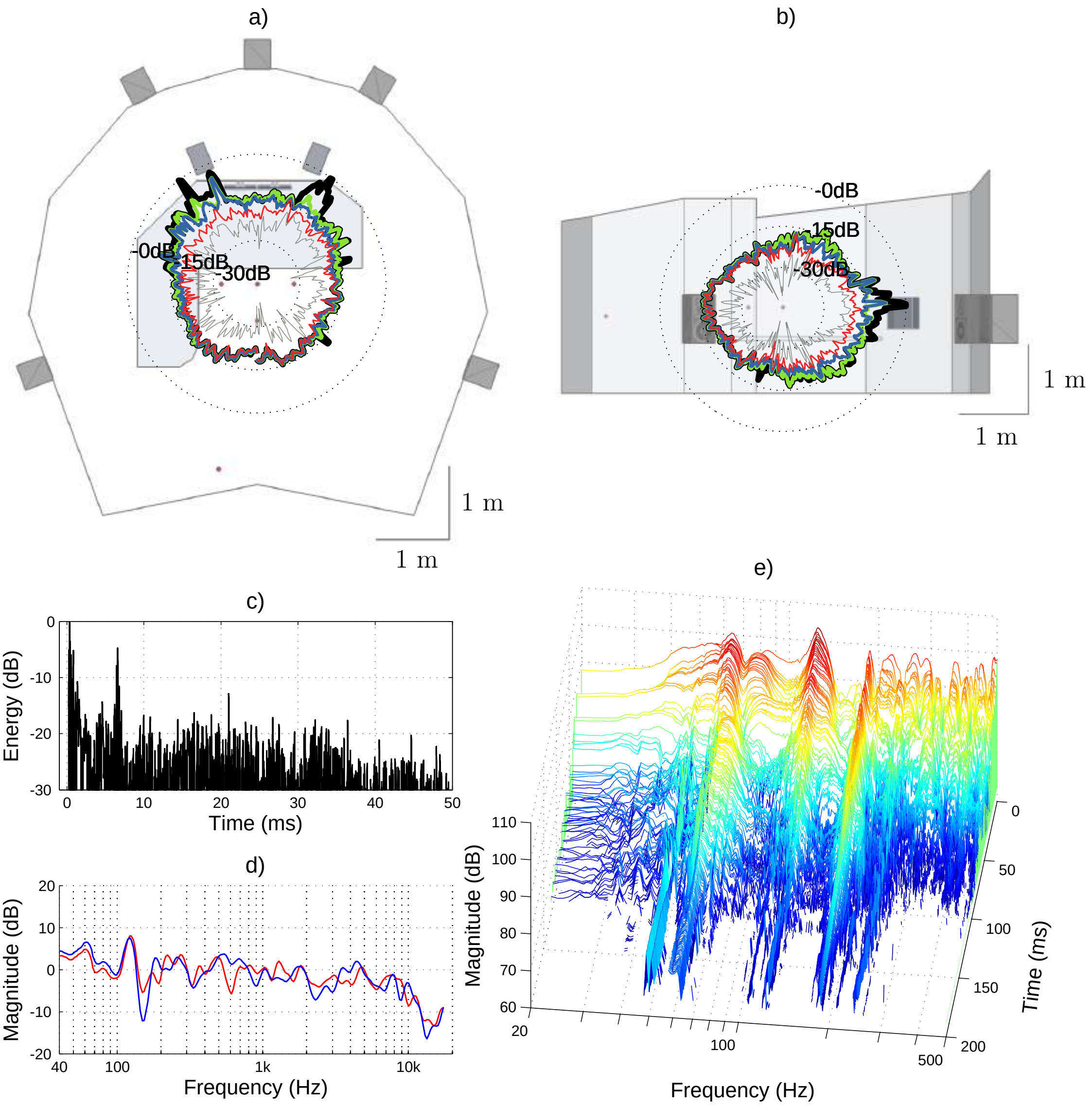


Figure 36: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#8.

A.9 CR#9

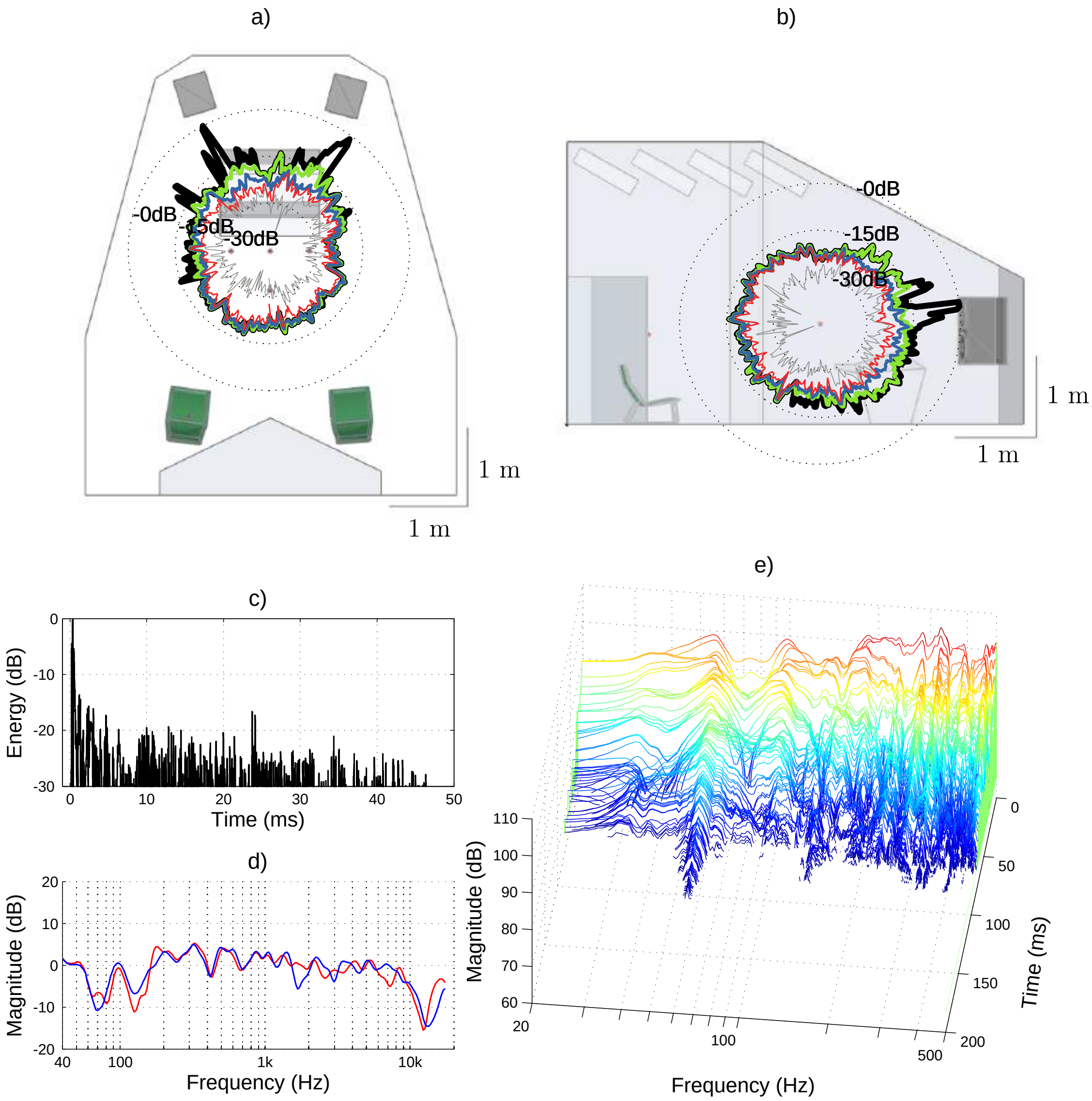


Figure 37: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#9.

A.10 CR#10

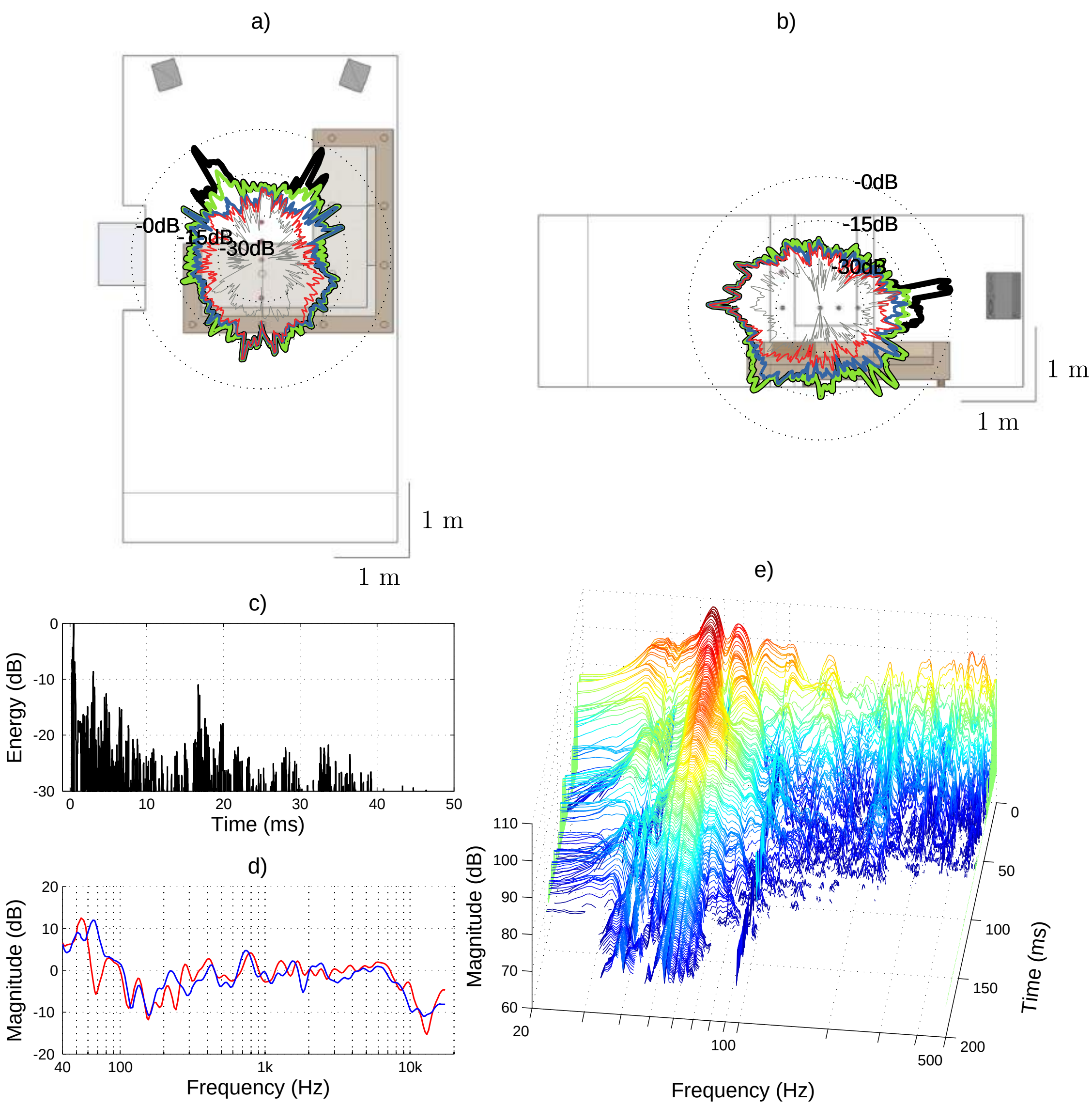


Figure 38: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#10.

A.11 CR#11

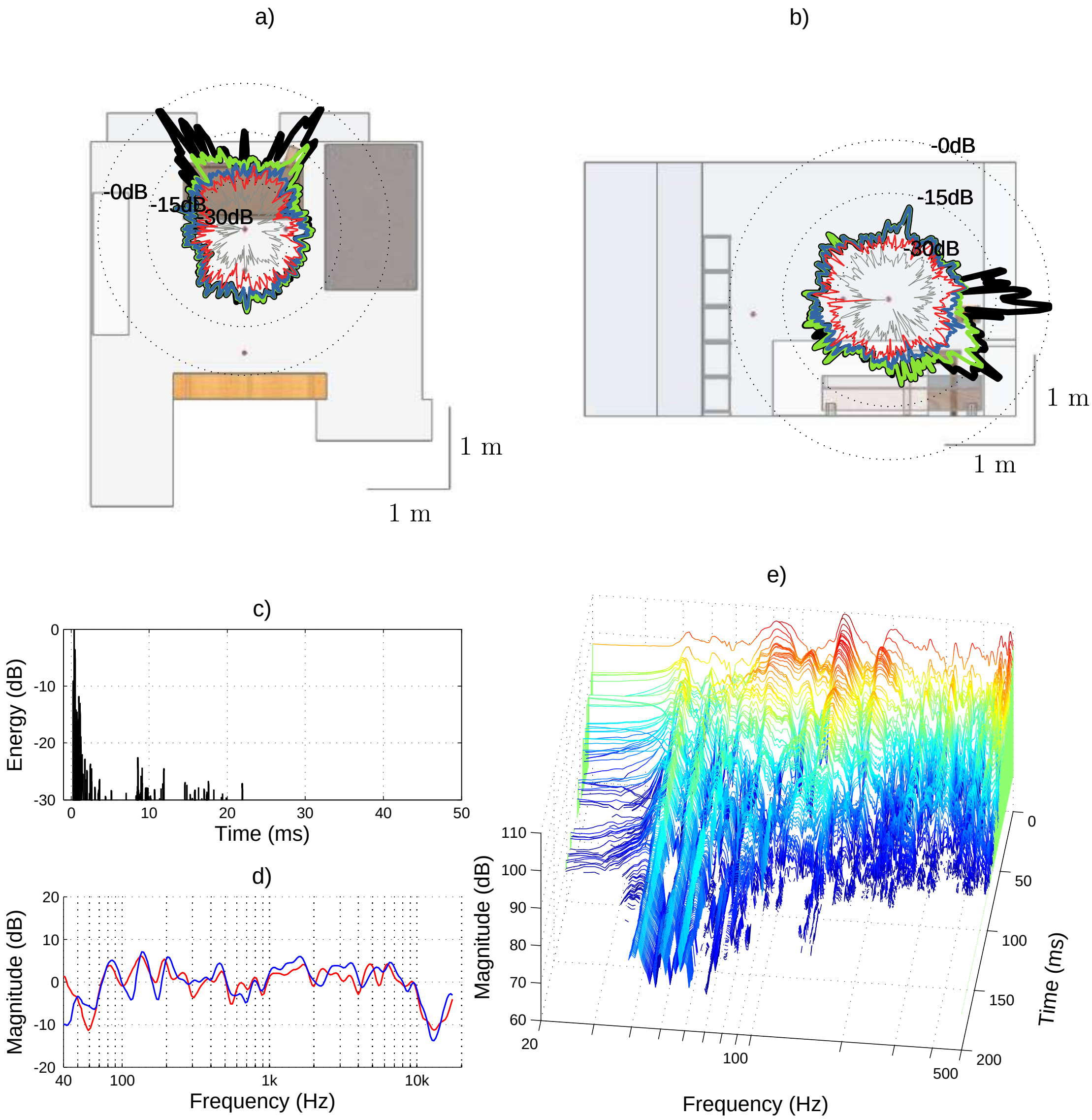


Figure 39: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#11.

A.12 CR#12

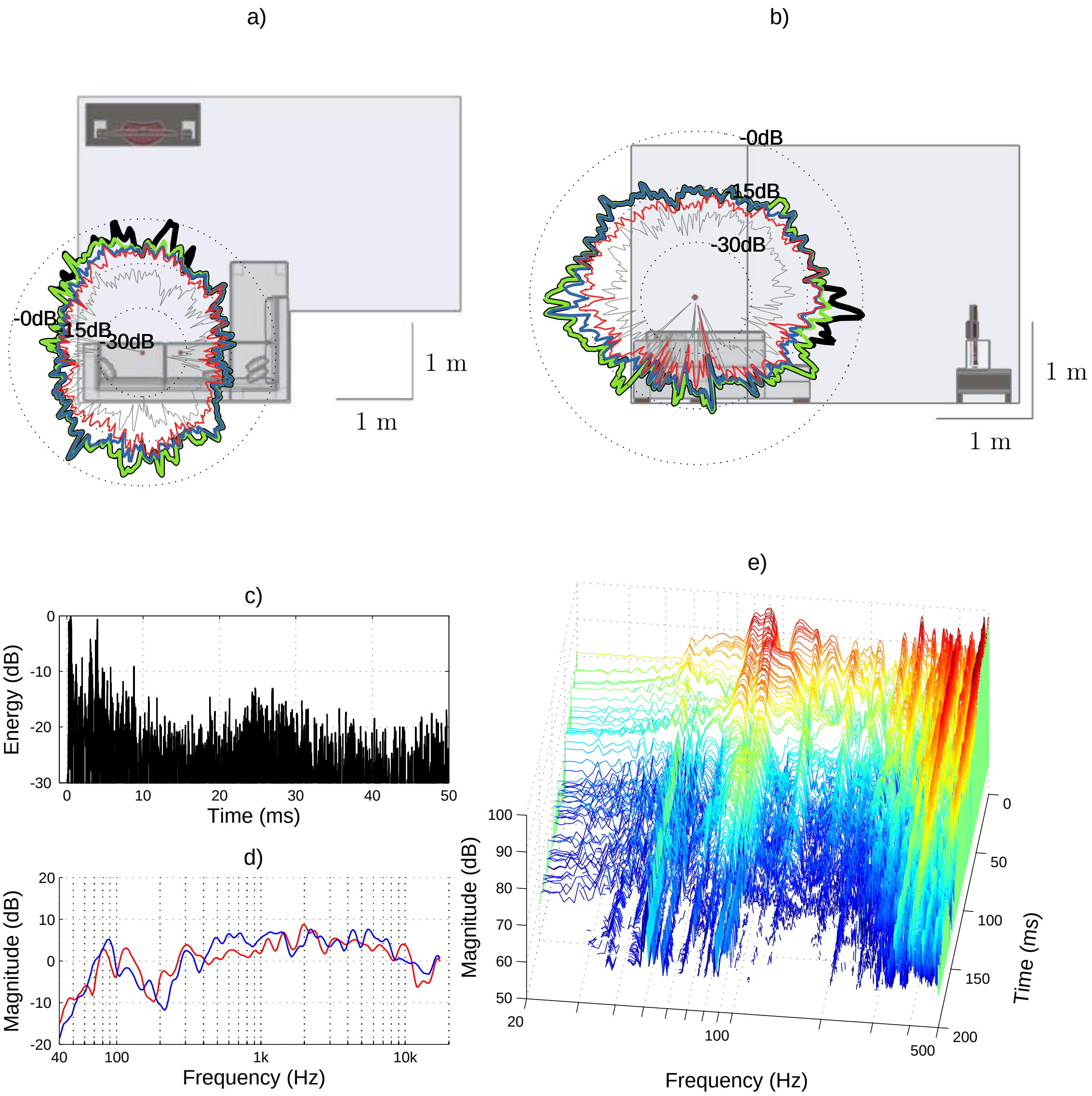


Figure 40: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#12.

A.13 CR#13

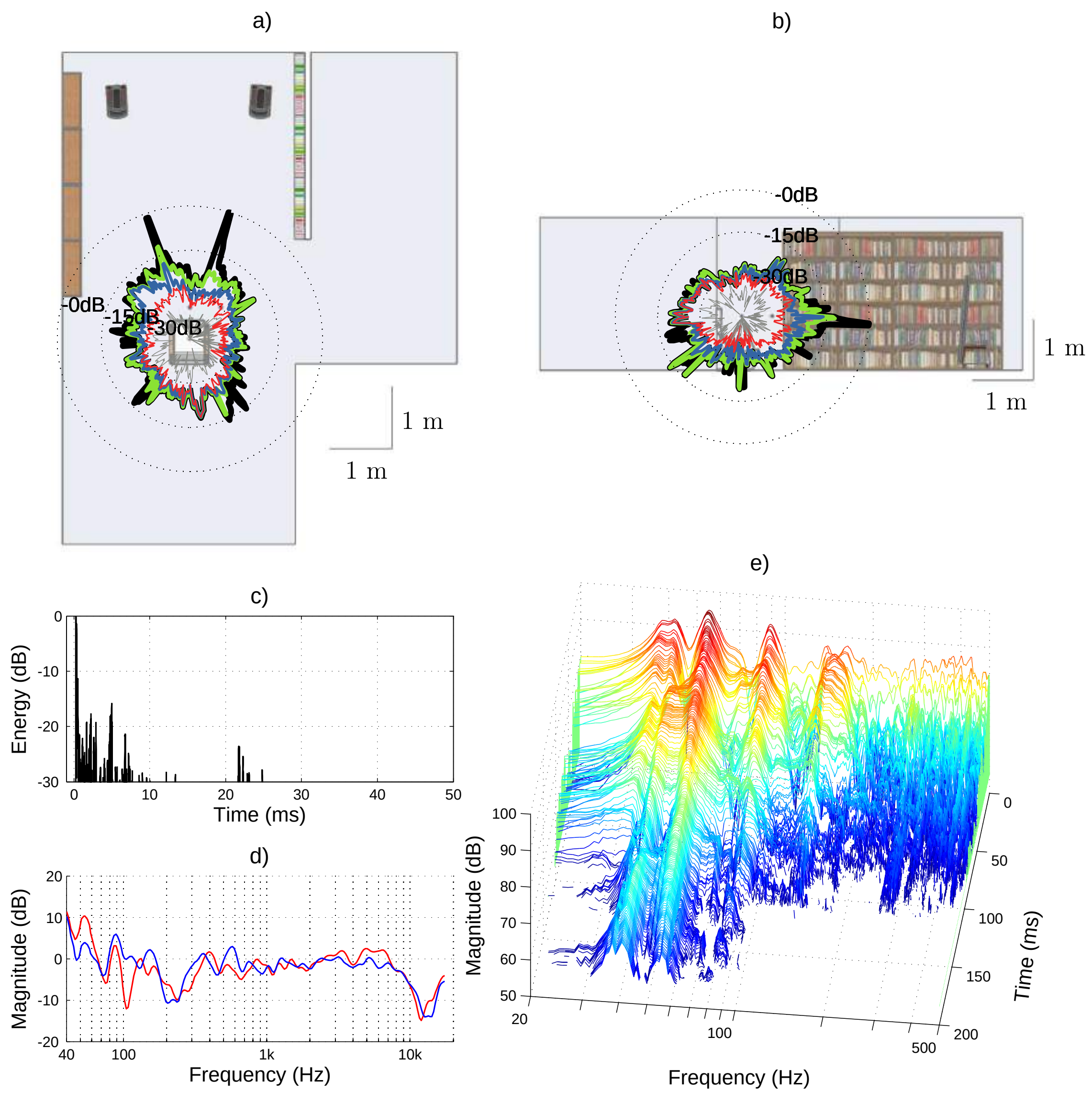
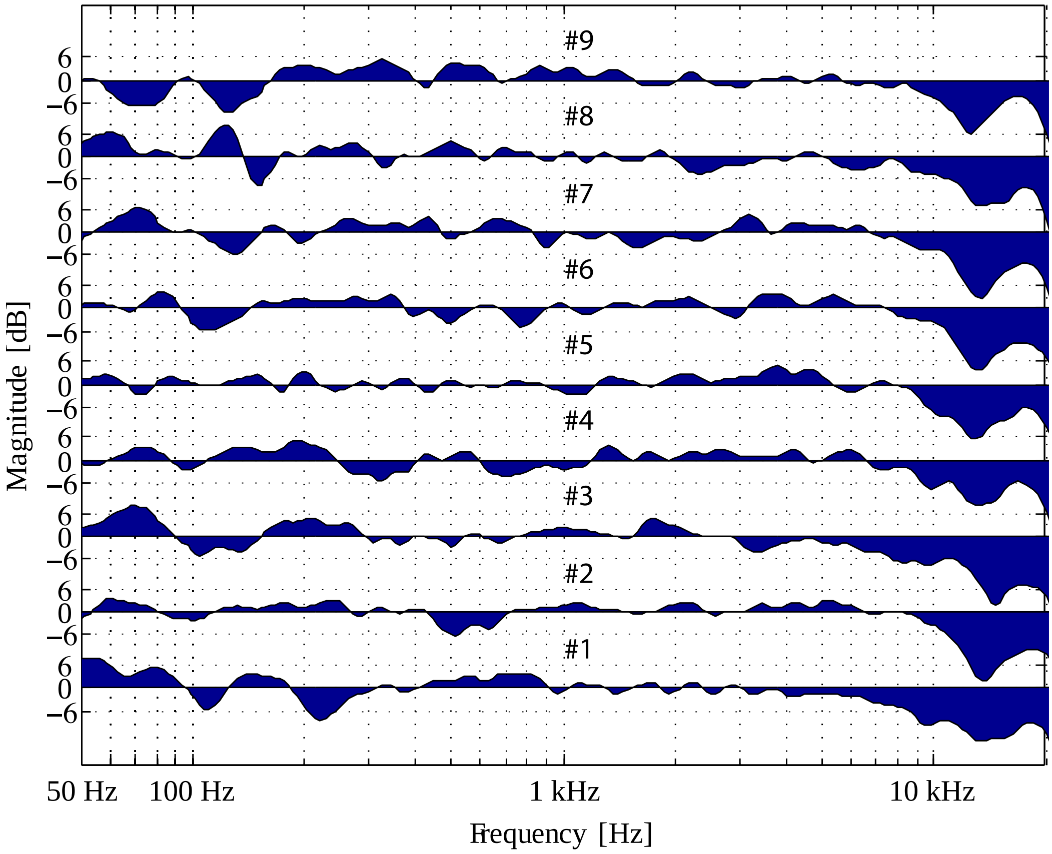


Figure 41: a) Cumulative polar response in lateral plane, b) Cumulative polar response in median plane, c) ETC, d) Magnitude response and e) CSD (waterfall) of CR#13.

B Magnitude response comparison (R1)

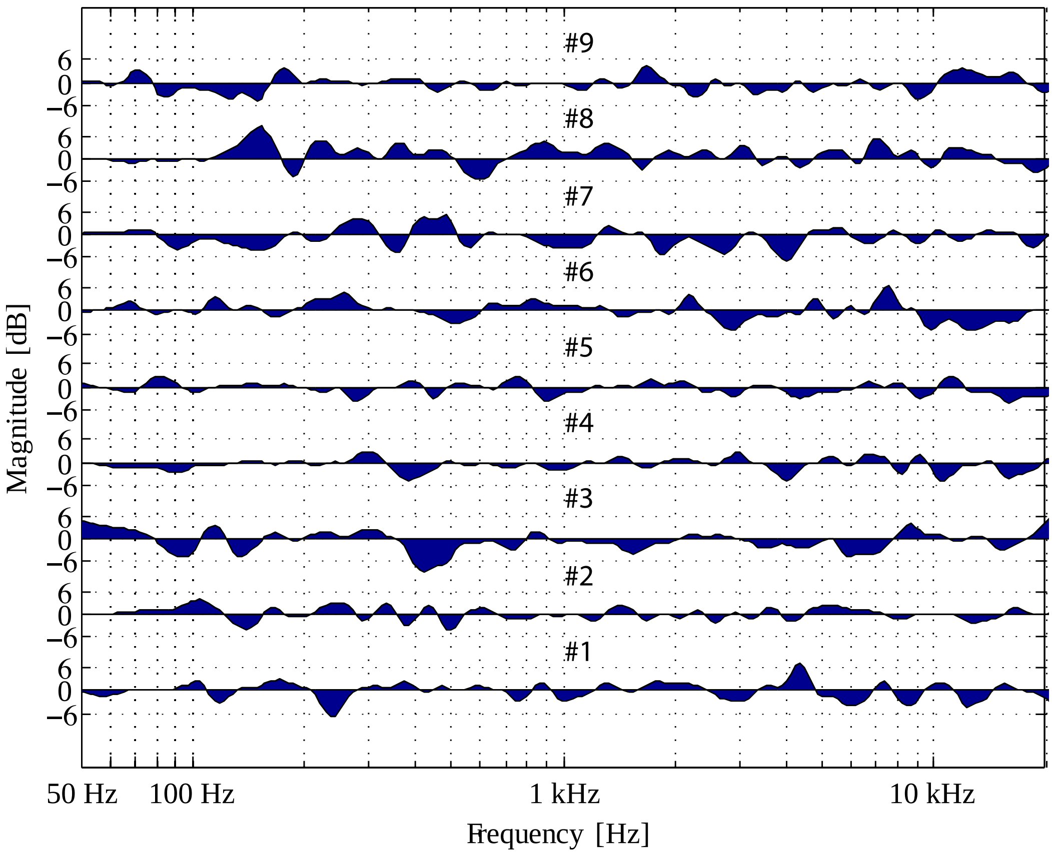
B.1 Magnitude responses of different control rooms



CR #	Loudspeakers	Purpose
1	3W-p	MA
2	2W-a	C/MI
3	3W-p	C/MI
4	3W-a	MA
5	2W-a	C/MI/T
6	3W-a	C/MI/T
7	2W-a	C/MI
8	3W-a	C/MI/T
9	3W-p	MA

Figure 42: Magnitude responses of different control rooms. Every magnitude response is normalized in a way that 0 dB corresponds the average magnitude between 50 Hz and 16 kHz. a: Active loudspeaker, p: Passive loudspeaker, 2W: 2-way loudspeaker, 3W: 3-way loudspeaker C: Control room, MI: Mixing room, T: Teaching room, MA: Mastering room

B.2 Magnitude response difference between left and right channel among control rooms



CR #	Loudspeakers	Purpose
1	3W-p	MA
2	2W-a	C/MI
3	3W-p	C/MI
4	3W-a	MA
5	2W-a	C/MI/T
6	3W-a	C/MI/T
7	2W-a	C/MI
8	3W-a	C/MI/T
9	3W-p	MA

Figure 43: Difference in magnitude response between left and right stereo channels among different control rooms. Difference is calculated by deducing a left channel from a right. Every magnitude response difference is normalized in a way that 0 dB corresponds the average between 50 Hz and 16 kHz. a: Active loudspeaker, p: Passive loudspeaker, 2W: 2-way loudspeaker, 3W: 3-way loudspeaker C: Control room, MI: Mixing room, T: Teaching room, MA: Mastering room