Internship report

March 2017 - August 2017

# Influence of Room acoustics on the performance of a musical ensemble

*Author:*
Vincent MARTIN

*Supervisors :*
Zora Schaerer Kalkandjiev &
Stefan Weinzierl

Technical University of Berlin, Audio Communication Group

# Abstract

This report aims at establishing, in an empirical way, the influence of the acoustical characteristics of a sound environment on the musical performance of a group of musicians. The environment acts as a transformer of the sound produced by a musician, and the adjustments during a performance are realized intuitively by the latter. Some musical treatises of the 18th and 19th centuries are dedicated to the musical techniques that can be used in order to adapt to specific acoustical conditions. For instance, it is generally assumed that high reverberation imply a slower tempo.

However, the specific relation between musical performance and room acoustics was rarely the object of scientific research. The work accomplished during my internship is a continuation of the work carried out by my two tutors at the department "AudioKommunikation" of the Technical University of Berlin : Pr. Stefan Weinzierl and Dr. Zora Schaerer-Kalkandjiev. The first conducted studies to describe the performance of musicians using descriptors taken from recordings of various musical performances (solo, chamber music, grand orchestra) while the second worked during her thesis on stage acoustics' coefficients that are most influential during a musician's solo musical performance. Those works will be the basis of the state of the art in this report.

A database of recordings was already available at the beginning of the internship: several quartets of musicians were recorded in different rooms. Initially, a method of onset detection adapted to the database, exploiting in particular the score of the played songs, is developed. The descriptors of the performance are then extracted using the division of the recordings previously carried out. Then, based on the acoustic characteristics of the different rooms a statistical model is developed establishing a relation between these acoustic characteristics and those of the different performances recorded.


**Keywords** : Musical performance, Sound environment, Onset detection, Performance descriptors

# Resumé

Ce rapport de fin de stage vise à établir, de manière empirique, l'influence des caractéristiques acoustiques d'un environnement sonore sur la performance musicale d'un ensemble de musiciens. L'environnement agit comme un transformateur du son produit par un musicien, et les ajustements dans la performance sont réalisés de manière intuitive par ce dernier. Ils existent certains ouvrages du 18ème et 19ème siècle qui traitent des techniques de jeu recommandées pour s'adapter à certaines conditions acoustiques. Par exemple il est souvent admis qu'une réverbération importante implique de jouer avec un tempo plus lent.

Cependant la relation entre performance musicale et acoustique des salles fut rarement l'objet de recherche scientifique. Le travail réalisé pendant mon stage s'inscrit dans la continuité des travaux réalisés par mes deux tuteurs au département "AudioKommunikation" de l'Université Technique de Berlin Pr. Stefan Weinzierl et Dr. Zora Schaerer-Kalkandjiev. Le premier a conduit des études visant à décrire la performance de musiciens à l'aide de descripteurs extrait d'enregistrements de performances musicales diverses (solo, musique de chambre, grand orchestre) tandis que la seconde a travaillé au cours de sa thèse sur les coefficients d'acoustique des scènes qui sont les plus influents au cours de performance musicale solo de musicien. Cela constituera la base de l'état de l'art de ce rapport.

Une base de donnée d'enregistrements était déjà disponible au début du stage : plusieurs quartets de musiciens ont été enregistrés au sein de différentes salles. Dans un premier temps une méthode de détection d'attaques adaptée à la base de données, exploitant nottament la partition des morceaux joués, est développée. Les descripteurs de performance sont ensuite extraits en fonction du découpage des enregistrement précédemment réalisé. Ensuite à partir des caractéristiques acoustiques des différentes salles un modèle statistique est développé pour établir une relation entre ces caractéristiques acoustiques et celles des différentes performances enregistrées.

**Mots-clés** :Performance musicale, Environnement sonore, Détection d'attaque, Descripteurs de performance

## Acknowledgements

I first would like to thank Stefan Weinzierl and Zora Schaerer-Kalkandjiev for welcoming me on the campus of the Technical University of Berlin and accepting me among their teams. I particularly thanks Dr. Schaerer for the supervision of my Internship, for her support and enthusiasm in this project, and for all the knowledge she shared with me.

I would also like to thank all the members of the Audio Communication Group for their warm welcome.

Finally, sincere thanks to the ATIAM supervision team and all my fellow students for this amazing year we shared.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 The AudioKommunikation Group

This internship took place in the "AudioKommunikation Group" of the Technical University of Berlin (TU-Berlin) which is one of the major scientific university in Germany. The AudioKommunikation Group focuses on research and teaching in communication of music and speech in acoustical or electro-acoustical systems. In particular the projects conducted at the department relate to the reproduction of sound environment with binaural synthesis or sound field synthesis, the study of the musical content,and technologies for composition and realization of electro-acoustic music. The department also organizes a master program in Audio Communication and Technology.

The department, leaded by Pr.Weinzierl, runs two electronical studios with a 12 resp. 8 channel loudspeaker setup, a wave field synthesis laboratory with 192 channels, a 3D media lab including 180° panorama projection and dynamic binaural reproduction, and the world's largest wave field synthesis installation.

My two supervisors Pr. Weinzierl and Dr. Schaerer Kalkandjiev run a project funded by the German Research Foundation (DFG) "Room acoustics and the performance of music" which aims to investigate the influence of room acoustics on solo and ensemble music performances. The work done during this internship represent my contribution to this project.

## 1.2 Objectives

As mentioned, the work I have done during my internship is part of a larger project which aims to identify the influence of room acoustics on music performances. This project will only focus on small ensembles of musicians (string quartet and brass quartet) playing classical pieces. As the use of empirical method is what was chosen at the beginning , this results to divide the problem in three different research focuses :

In one hand it was needed to investigate on how to characterize the performance of an ensemble with a set of coefficients, following a method that does not need a human intervention.

In the other hand, it was important, before any statistical analysis, to choose which acoustical properties were relevant to the point of view of a musician and extract the corresponding coefficients.

In the end, based on the results of these two previous investigations, it will be possible to test different regression models in order to enlighten the relation between room acoustics and performance characteristics.

## 1.3 Organization of the report

Before the analysis of the recordings, it was important to study the early theory about performance and room acoustics. Therefore, **Chapter 2** will focus on the state of the art and some prerequisites for the comprehension of stage acoustics and the software associated that was used

**Chapter 3** will describe how the recordings of the ensembles were made, and the nature of them.

**Chapter 4** will focus on the method used for the analysis of these recordings, the goal during this part of the work was to be able to characterize the performance of each recording

**Chapter 5** will develop the statistical method used to link the characteristics obtained in the previous chapter and the acoustical coefficients of the concert venues corresponding to the recordings and how those where selected.

Finally, **Chapter 6** will present the results obtained through this empirical investigation and will be evaluated, the conclusion and the possible further research and improvements will be in **Chapter 7**

# Chapter 2

# State of the art

## 2.1 Performance and room acoustics

In this section, the concept of performance will be defined as it is considered in this report. Moreover, the litterature as the point of view of musical scholars and, more recently, room acoustic research will be listed below.

### 2.1.1 The concept of performance

A musical performance can be defined as a unique physical rendition of a score into a listener's experience. This concept implies that a performance is not only defined by the score, but also by the interpretation of the performer himself, According to Clarke [4] : "animate the music, to go beyond what is explicitly provided by the notation or aurally transmitted standard - to be 'expressive'". However a performance is supposed to represent the musical ideas of the composer, for a better understanding Figure 1 propose a representation of the performance process.

In this chain of communication the room acoustics play the role of a transformer of the sound event which will influence the listener's and performer's experience. That will lead the performer to correct his way of playing in order to adapt specifically to his acoustic environment.

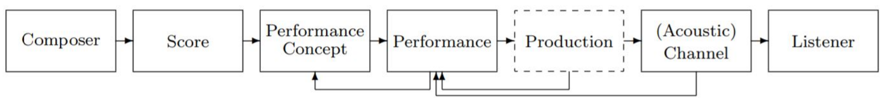


Figure 2.1: Chain of communication for a music performance (from [1])

### 2.1.2 Musical scholars' point of view

Scholars and musicians produced recommendations on the way to adjust depending of the room acoustics, several musical treatises of the 18th and 19th century explained precisely how a musician should react to specific acoustical conditions.

For instance, it was admitted that a larger room implies that a musician should play on a slightly slower tempo [11] "If one plays in a big place that resounds strongly a slightly slow trill will have a better effect than a fast one. Because an all too quick movement of the tones is confused by the reverberation and hence the fast trill is blurred." This relation between room volume, reverberation time and certain temporal aspects in a musical performance is still up to date today. The violonists Borciani and Galamian [9] also theorized other aspects of the musical performance, for solo and quartet performances, such as dynamics and articulation.

To sum up with Flesch's and Blum's work : [6] [2]

- Slower tempo in case of large reverberation.
- Weaker loudness in large halls, wider vibrato, separated ligatures. In dry halls, on the contrary, a prolongation of tones is recommended.
- Weaker attack in case of a large hall.
- In large halls but with small reverberation larger loudness.

### 2.1.3 Room acoustic research

Most of the literature focuses on the influence of room acoustics on the tempo and dynamic strength (loudness).

Naylor [18] studied the perception of tempo by musicians in different rooms, in his experiment different musicians had to tap synchronised with the tempo of different pieces. The decay of the notes were artificially intensified in some cases to simulate reverberant and dry rooms. Naylor brought to light that the error of synchronisation were increasing with a higher decay as the attack of the tones were less defined when musicians hear it. Moreover, in real performing conditions he observed that sharpening the attack was a way for musicians to "adapt" to high reverberant rooms.

Kawai et al. in 2015 and Bolzinger S. in 1994 [13][3] measured different musicians playing a grand MIDI piano for different pieces in different simulated sound fields producted with loudspeakers. The tempo, loudness and use of the pedal were measured. And if the results for tempo were unsystematic, loudness appeared to be negatively correlated to the quantity of early sound reflections (section 2.2), and a very strong negative correlation was brought to light between the reverberation time (section 2.2) and the use of the pedal which was intuitively described in musical treatises.

Kato et al. (2007) [12] and Schaerer (2015) [22] pointed similar correlations, and through interviews of musicians have shown that most performers are conscious of these effects and that tempo was not taken into account in their way of adapting to different room acoustics.

The different studies agrees on these different aspects of the influence of room acoustics on musical performances (solo and ensemble) :

- Loudness is increased in rooms with small reverberation time
- Higher reverberation time led to the use of sharpened attacks (stronger agogic)
- No clear relation on tempo, the influence of the type of instrument seems to be high and the technique of adapting may differs from a musician to another

These results incite to explore other performance and room acoustical parameters, and not only those quantifying reverberation of a room as they do not seem to give enough visibility to establish relations. That is why other room acoustical parameters will be taken into account in this report. (Section 2.2)

## 2.2 Stage acoustics theory

### 2.2.1 Stage acoustics coefficients

Considering the previous studies, it seems important to cover more aspects of room acoustics than just reverberation, as musicians are able to differentiate more aspects. And being able to differentiate these different aspects mean that there is a possibility that these aspects influence their way of playing.

Gade (1986) [8] ran a series of interview to list all the aspects of room acoustics perceived by musicians, which was completed by Sanders (2003) [20] : 'reverberance', 'support', 'timbre', 'dynamics', 'clarity', 'balance' and 'warmth'. These research contributed to establish the ISO 3382-1 (2009) which standardizes the measurement parameters of room acoustics. The following paragraphs contain the different coefficients used in the further analysis of acoustic environments.

Furthermore, most of these coefficients will be calculated for third octave bands. To reduce the number of value per coefficients, an average will be done over certain third-octave band as it is recommended for the measurement of room acoustics in ISO 3382-1. (Annex A for the details)

#### Reverberance

Reverberation time $RT$ is the time for a sound to decay, between a source and a receiver, of a certain amount of decibels, after the perception by the receiver of the direct sound of the source is not received. In this internship we will use the value of 30 decibels.

Early decay time $EDT$ is the decay time for 10 decibels. This parameter correspond to the perception of the strength of the reverberance, while $RT$ is considered to be for the perception of the duration of the reverberation.

**Support**

According to performers, the important factors during a performance for a musician in an orchestra is the capacity to hear his own instrument, the others musicians' instruments, and finally the acoustic response of the hall. [7] These different sounds can be considered in three categories : the early reflections, the direct sound, and the late reflections.

The most recognized parameters to quantify these reflections in stage acoustics are those proposed by Gade. [8] These two parameters have been integrated in the standard ISO-3382-1.

For describing the assistance of early reflections the parameter used is $ST_{early}$:

$$ST_{early} = 10log(\frac{\int_{20ms}^{100ms} p_{1m}^2(t)\,\mathrm{d}t}{\int_{0ms}^{10ms} p_{1m}^2(t)\,\mathrm{d}t})[dB]$$

$p_{1m}$ : the pressure measured at 1m distance from the source

It quantifies the proportion of energy of the early reflections compared to the direct sound.

To describe the perception of the reverberation of the hall the parameter used is $ST_{late}$ :

$$ST_{late} = 10log(\frac{\int_{100ms}^{1000ms} p_{1m}^2(t)\,\mathrm{d}t}{\int_{0ms}^{10ms} p_{1m}^2(t)\,\mathrm{d}t})[dB]$$

The upper limit of 1s was chosen to save calculation time considering the influence of later response as neglictable in the perception of the acoustic of most halls.[27]

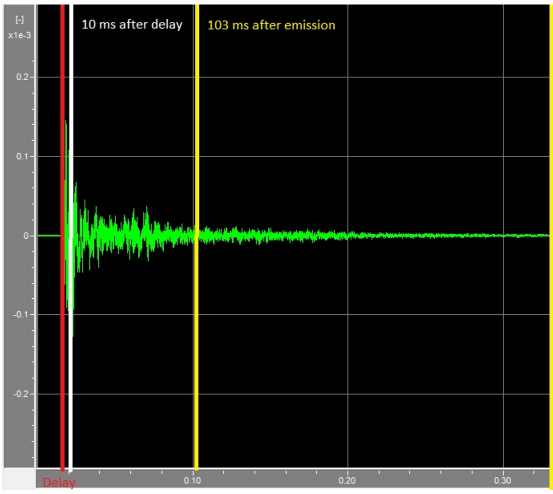


Figure 2.2: Example of an impulse response

In figure 1 : the part between the red line (delay of the sound between the source and the receiver) and white line (10ms after delay) is the part considered as the direct sound, between the red and yellow line (103 ms after delay) is the part corresponding to the early reflections, the rest after the yellow line corresponds to the late reflections.

These parameters have been extended by R.Wenmaekers et al.[28] in order to be used in different source-receivers distance $d$ configurations by integrating the delay time :

$$ST_{early,d} = 10log(\frac{\int_{10ms}^{103ms-delay} p_d^2(t)\,\mathrm{d}t}{\int_{0ms}^{10ms} p_{1m}^2(t)\,\mathrm{d}t})[dB]$$

$$ST_{late,d} = 10log(\frac{\int_{103ms-delay}^{infinity} p_d^2(t)\,\mathrm{d}t}{\int_{0ms}^{10ms} p_{1m}^2(t)\,\mathrm{d}t})[dB]$$

These two lasts coefficients will be used to evaluate the 'support' of a room for a musician over the previous ones.

### Clarity

The clarity of sound is defined by the logarithmic ratio of the energy of the early sound by the late sound :

$$C_{t_e} = 10log(\frac{\int_{0ms}^{t_e} p^2(t)\,\mathrm{d}t}{\int_{t_e}^{infinity} p^2(t)\,\mathrm{d}t})[dB]$$

ISO 3882-1 advises a value of $t_e = 80ms$, which will be used in the following, the clarity coefficient will be noted $C_{80}$.

**Dynamics**

The loudness of the sound is measured by the sound strength coefficient $G$ which is the logarithmic ratio of the sound energy (square and integration of the sound pressure) of an impulse response measured in the room to that of the response measured in a free field at a 10 meters distance.

$$G = 10log(\frac{\int_0^{infinity} p^2(t)\,\mathrm{d}t}{\int_{0ms}^{infinity} p_{10m}^2(t)\,\mathrm{d}t})[dB]$$

Moreover, the coefficient $G125$ representing the sound strength for the octave band 125 Hz will be used.

Two alternatives to the support parameters have also been proposed by Dammerud [5] considering that the 1 meter distance between source and receiver used in them could cause errors because of the directivity of the source.

He introduced the early strength coefficient $G_e$

$$G_e = 10log(\frac{\int_0^{80ms} p^2(t)\,\mathrm{d}t}{\int_{0ms}^{infinity} p_{10m}^2(t)\,\mathrm{d}t}) = 10log(\frac{10^{C_{80}/10} * 10^{G/10}}{1 + 10^{C_{80}/10}})[dB]$$

And the late strength coefficient :

$$G_l = 10log(\frac{\int_{80ms}^{infinity} p^2(t)\,\mathrm{d}t}{\int_{0ms}^{infinity} p_{10m}^2(t)\,\mathrm{d}t}) = 10log(\frac{10^{G/10}}{1 + 10^{C_{80}/10}})[dB]$$

**Warmth**

The coefficient used for defining the 'warmth' of a hall is the bass ratio $BR$, according to ISO 3382-1 :

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

### 2.2.2 Ray-tracing and image sources simulation

The acoustics of a room can be modeled with different methods and for different purposes. There are two main approaches : one is based on the numerical solving of the wave equation by a finite element method for example. The other method is based on the assumption of geometrical acoustics. *Raven* [19] [1] [25] is based on this second approach. An architectural model, with surfaces in which scattering and absorbing coefficients have been assigned, acts as an input. Geometrical acoustics considers the propagation of sound as rays, in such a model an omnidirectional source emits rays in random directions. Each ray possess a fixed energy which is lost through the air and when it bounces on absorbent surfaces. The receiver is modeled as a small volume that build an impulse response from the rays it received.

In fact this ray-tracing techniques is only used for late reflections, for early reflections *Raven* uses the image sources technique. For a certain number of the nearest surfaces of a chosen receiver, image sources are placed symmetrically in relation to the surface (figure 2) and integrated to the model. Only first order image source are used. The early part of the impulse response is then calculated with the receiver and the several sources.

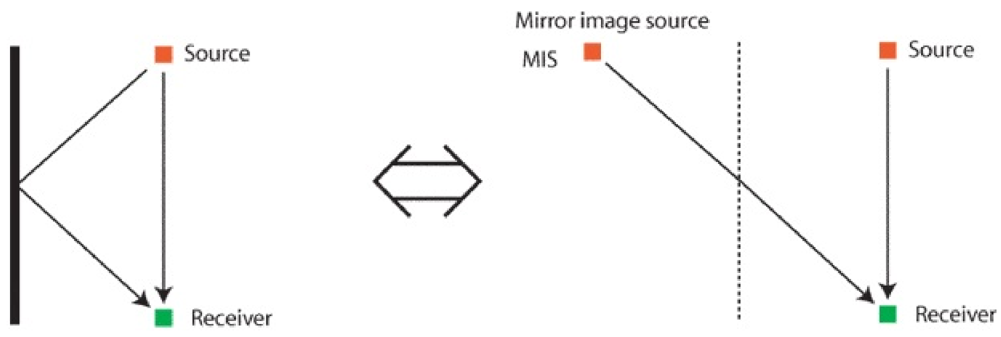


Figure 2.3: First order image source

It is a software used for complex room modeling, considering the computational time, geometrical acoustics is a lot more shorter than with a numerical solving. The assumption of ray-tracing is valid in high frequencies in which the wave length of sound is short considering the size of surfaces usually present in room acoustics modeling. However at low frequencies the approximation error increases as wave phenomena are not neglictable, especially if there is obstacles on the path of the direct sound [21], which should not be the case in this study.

The support parameters were not integrated in the default code of *Raven* an additional function was added to the software before the extraction of the rooms' acoustic parameters. (Appendix C)

# Chapter 3

# Recordings

The recordings were already available at the beginning of the internship. They were recorded in the studio of the AudioKommunikation group using headphones and a binaural reproduction technique to reproduce the acoustical environment. For each room, the impulse response between each source and receiver (instrument and musician) was previously calculated with *Raven* and the signals from microphones put on the instruments convoluted in real time, so each musician could hear himself and the others as he was in the real room. [17]

2 different quartets (1 brass quartet, 1 string quartet (Figure 3.1)) were playing 3 different pieces in 14 virtual acoustic environments, each instrument were recorded separately in a 4-track .WAV file. This will constitute the database of the project.

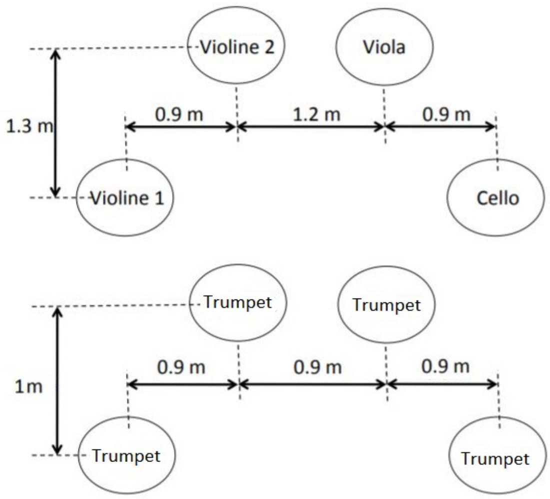


Figure 3.1: Quartet configurations

## 3.1 Concert venues

Among the 14 different rooms the goal was to provide the largest spectre of configurations possible and still being likely used for a chamber music concert.

To illustrate the diversity figure 3.2 shows the evolutions of different coefficients representing various aspects of the rooms simulated.

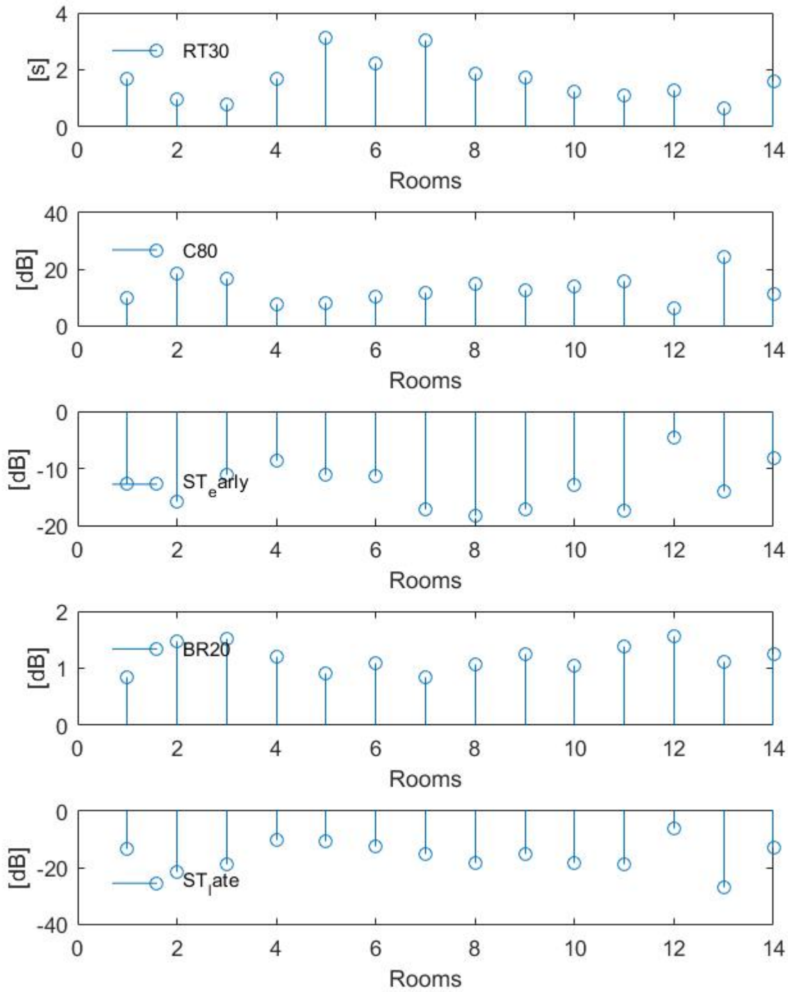


Figure 3.2: Average of the acoustic coefficient over every source-receiver (musician-instrument) values calculated via *Raven*

## 3.2 Pieces

The pieces played were :

- For the brass quartet : "Fantasia & Fugue in G Minor for Woodwind Quartet" by Bach, arrangement by Magatagan M.
- For the string quartet : "String Quartet in D major, Op. 76 No. 5, Hob. III:79" by Haydn

The advantages of these pieces is that they can be both considered as romantic chamber music, so the results of the study by Weinzierl & Maempel which explains how to extract performance characteristics from audio features (for romantic-classical music only). These performance characteristics will be needed afterwards.

# Chapter 4

# Performance analysis

In this Chapter, the method used to analyse a performance, by mesuring different physical parameters, in the recordings will be explained. This work will be mainly based on an investigation by Weinzierl & Maempel (2011) , and by using a feature extraction method developed by Lerch (2009). [15]

In a first time, a detection of the onsets (3.1) will produce a grid for each recordings, then a repertoire of audio features will be extracted following this grid (3.2). Finally, these features will be used to calculate the different performances characteristics. (3.3)

## 4.1 Onset Detection

Processing an onset detection on the recordings was crucial in order to define, not only, a grid dividing the recordings in intervals for extraction of audio features, but also to extract informations about the tempo which are among the most important parameters.

There are several method described in the literature for automatic tempo extraction/onset detection :

#### Blind method

The majority of the published algorithms can be defined as "blind" methods, as it doesn't need any information on the digital audio content, it only need the audio data as input. For instance Scheirer [23] proposed a tempo extraction system based on a bank of resonance filters. By using the derivative of the audio signal's envelope as input of these filters, the comb filter corresponding to the tempo will have the biggest output, and define a single tempo for the audio data. Klapuri established a similar approach. [14]

These methods are usually not adapted to sudden tempo changes and score containing several silences, which is the case in this project.

**Performance-to-score matching**

In this project there was the opportunity to know which score was played for each recordings, so additional informations in form of MIDI scores could be used to improve the accuracy of the onset detection. That is why approaches called performance-to-score matching were used. More specifically the method proposed by Lerch (2009) [15] was used, and his software coded on Matlab was improved by the use of *Ableton Live*.

### 4.1.1 Used algorithm (Lerch)

The figure 4.1 show the different steps of the software developed by Lerch on Matlab, with an audio file and a midi file as inputs and the list of onsets for the corresponding recording as output:

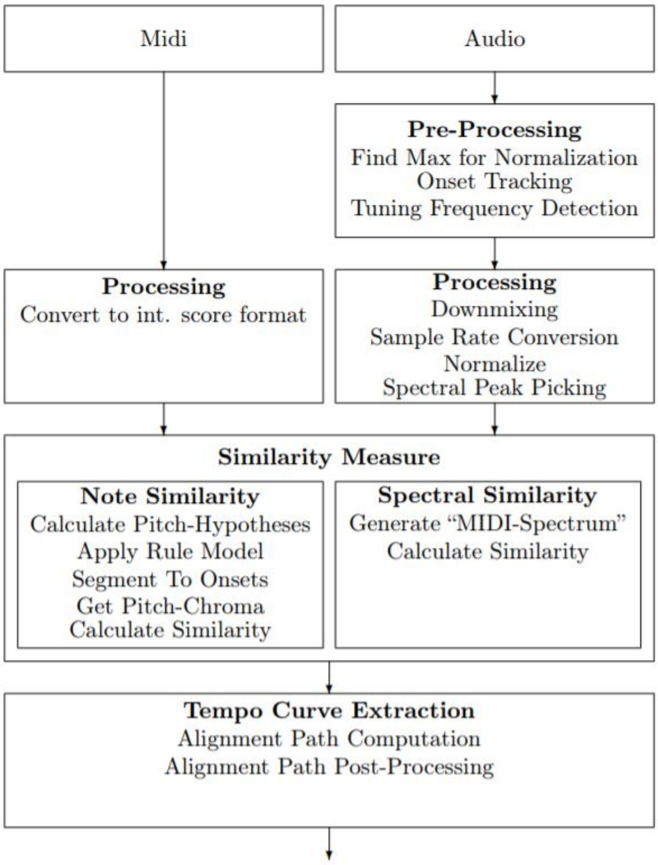


Figure 4.1: Flow chart of the processing steps for the tempo extracton stage (from [15])

The algorithm is divide between 3 main blocks per file :

**Pre-Processing and Processing**

The "¨Pre-Processing" and "Processing" blocks convert the midi file into an audio file, and reciprocally for the audio file.These blocks have been replaced by using *Ableton Live* which proposes intuitive conversion functions. It has indeed provided better results.

In Lerch's software in the conversion from audio to MIDI, the tonal content is extracted from the audio file, by calculating a Short Time Fourier Transform (STFT), and identifying 8 harmonics specific to the played instrument. An audio file is then synthesized on this assumption of the harmonic distribution. This model is too simple to fit the reality in this case, Lerch precises that his processing model is the best in case of a recording with several instruments because of the combination and variation of mixtures. So a more sophisticated model will not be of any help. However in this case every instrument are recorded separately so the use of the *Ableton Live* instrument's templates provide a better match with the real harmonic distribution.

Likewise, the audio to MIDI conversion was also performed by *Ableton Live* that have shown to be more accurate.

**Similarity Measure**

Two different similarity measures are used in Lerch's software :

- one from the midi converted to audio compared with the audio file (left part in figure 4.1) : *Spectral Similarity.*
- one from the audio file converted to a midi file (right part in figure 4.1) : *Note Similarity*

These measures are done by defining a distance between every observation of a file to every observation of the other in both similarity process.

This results in two similarity matrix, a weighted sum of both this similarity matrix is done, resulting in a matrix illustrating when the score and the audio file are similar. A path of high similarity from the upper left corner to the lower right corner of the matrix is calculated via a dynamic time warping algorithm, resulting in onset times for all the notes in the score.

### 4.1.2 Results

The detected onsets were checked by playing all the recordings with click sounds, more than 90% of it were correct, the rest was corrected with *Sonic Visualizer* a software that allow to visually check the onsets that need to be suppressed or added (Figure 4.2) with the help of the score.

The use of *Ableton Live* for the conversion gave a substantial increase of the rate of correct onsets from $< 70\%$ to $> 90\%$.

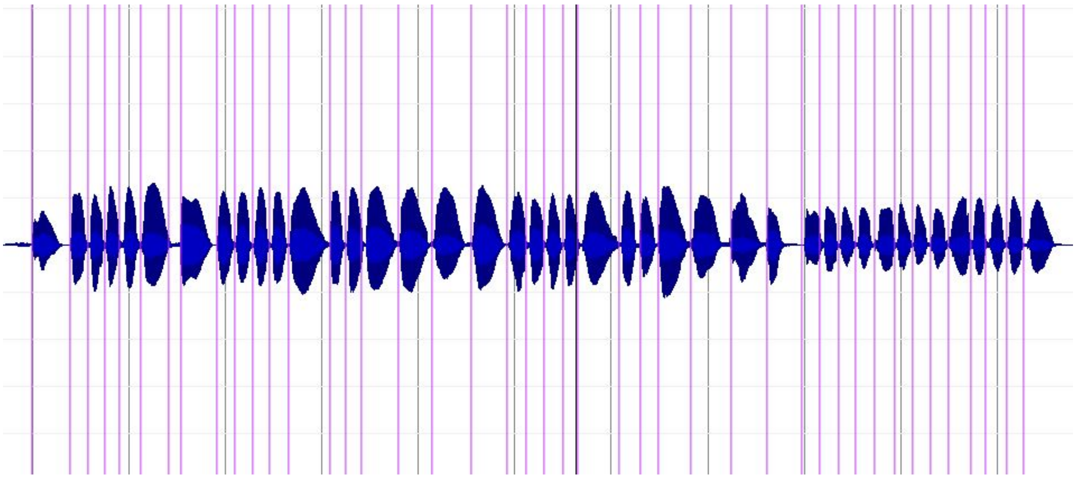

Figure 4.2: User interface of *Sonic Visualizer* (here the beginning of the first recording of one of the trumpet in the brass quartet)

## 4.2 Extraction of audio features

Once that all recordings were able to be separated following a grid of onsets the goal was to evaluate the performance of the musicians, with the help of audio features. So that these performance characteristics could be confronted to the acoustic parameters afterward. This part of the project will rely on the result of a study conducted by Weinzierl & Maempel at the AudioKommunikation Group in 2015. [24]

In this study a vocabulary was established by a group of different experts (musicologists, musicians, composers ...) suitable to characterize musician performances of a classic/romantic repertoire. (The vocabulary is in Appendix B)

Every characteristics in this vocabulary was evaluated by these experts for several performances, from these ratings a relation was established with a set of audio features presented below. The relation between these features and performance characteristics will be explained in section 4.3.

### 4.2.1 Audio features

In the following, will be presented the different audio features implanted in order to evaluate the performance characteristics of the recordings. Spectral and Loudness features algorithms were already coded. A more detailed description of these associated algorithms are present in Lerch's book. [16]

### 4.2.2 Time features

- inter-onset time, inter-bar time (time between two onsets or two bars)

$$IOI(i) = t(i+1) - t(i)[s]$$
$$IBI(k) = t(k+1) - t(k)[s]$$

$t(i)$ : Onset time of note $i$

$t_b(k)$ : Onset time of bar $k$

- With the use of the score, a normalization of the inter-onset time

$$IOI(i) = \frac{t(i+1) - t(i)}{\Delta\tau_{i,i+1}}$$

$\Delta\tau_{i,i+1}$ : Number of beats between two score events (one beat equal a quarter note)

- With the use of the score, the micro-tempo is calculated

$$TMPt_{note}(i) = \frac{60s}{t(i+1) - t(i)}.\Delta\tau_{i,i+1}[BeatsPerMinute]$$
$$TMP_{bar}(k) = \frac{60s}{t_b(k+1) - t_b(k)}.\delta\tau_{i,i+1}[BeatsPerMinute]$$

### 4.2.3 Loudness features

Several loudness measures can be used to extract dynamic features from an audio content. In the library proposed by Lerch. [16] A variety of algorithm have been implanted to predict the loudness of musical events on a local time scale, in the following are those used in Weinzierl & Maempel study and in this project :

- VU : The VU meter or volume indicator, it averages out fast level variations in order to better approximate perceived loudness. VU is calculated by smoothing the absolute value of the input sample with a second order low pass filter.
- dBA : A weighting function is applied before the measure of the instensity of the signal. This frequency-dependant function is supposed to reproduce human ear sensitivity to certain frequencies. (figure 4.3)
- ITU-R BS.1770 : An intensity-based loudness measurement that has been standardized in ITU (international telecommunication union) recommendation ITU-R BS.1770. Another weight function is used. (Figure 4.3)

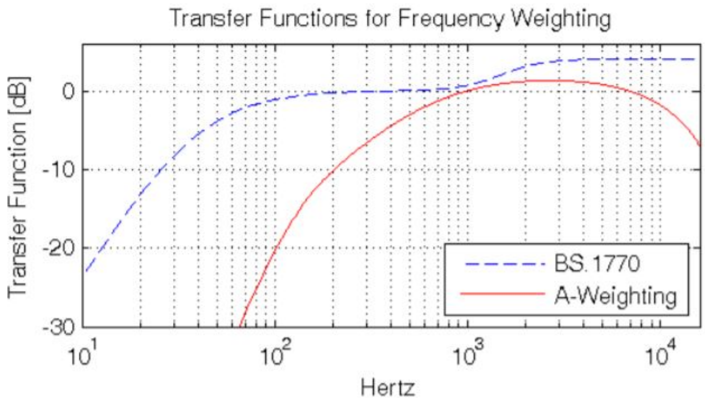


Figure 4.3: Frequency Weighting Transfer Functions (from Lerch (2012))

Based on the results of psycho-acoustic experiments, Zwicker [29] proposed a model for the measurement of loudness where masking phenomena are taken into account. Two approaches from the literature to represent Zwicker's loudness have been implanted :

- ZWDIN: An implementation of the Zwicker-Loudness according to DIN 45631:1991. (German Institute for Standardization)
- ZW1387: An implementation of the Zwicker-Loudness based on the psycho-acoustic model described in ITU-R BS.1387:2006.

### 4.2.4 Spectral features

The timbre of a sound, often also described as sound color, quality or texture, is considered as a third attribute of the perceptual experience of musical tones, besides pitch and loudness. Various attempts towards a definition of 'timbre' have been made. In contrast to loudness and pitch as unidimensional properties, timbre has been shown to be a multidimensional property, taking into account both spectral and temporal patterns. In the following are listed the spectral features necessary to complete the analysis of the timbre according to Weinzier & Maempel. All are based on the Short Time Fourier Transform (STFT) of the sample.

- SR: The Spectral Rolloff is a measure of the bandwidth of the spectrum. It computes the frequency below which 85% of the accumulated spectral power is concentrated.
- SF: The Spectral Flux measures the rate of change of the spectral shape
- SC: The Spectral Centroid is defined as the center of gravity of the power spectrum. The implementation follows the definition in ISO/IEC 15938:2002.
- SS: The Spectral Spread, sometimes also referred to as instantaneous bandwidth, describes how far the spectral power is spread around the SC. The implementation follows the definition in ISO/IEC 15938:2002.

- MFCC0-4: The implementation of the Mel Frequency Cepstral Coefficients (MFCCs) is based on the implementation in Slaney's Auditory Toolbox (Slaney, 1998).

## 4.3 Use of statistical descriptors

These different algorithms furnished a series of value for each audio feature. Based on the results of the regression models performed in the study of Weinzierl & Maempel, for each audio feature different statistical descriptors were calculated, a specific linear combination allowed to predict performance characteristic. (Results in Appendix B)

- Mean
- Geometrical Mean
- Median
- Mode (Element that occurs most often in the serie)
- Standard deviation (Square root of variance)
- Quantiles 10,90 (Value that seperate 10% / 90% of the lowest value to the rest)
- Interquantile distance 10-90

At this point of the project, the acoustical coefficients for each room are extracted, and the performance characteristics are evaluated for each recordings. (Illustrated in the graph figure 4.3)

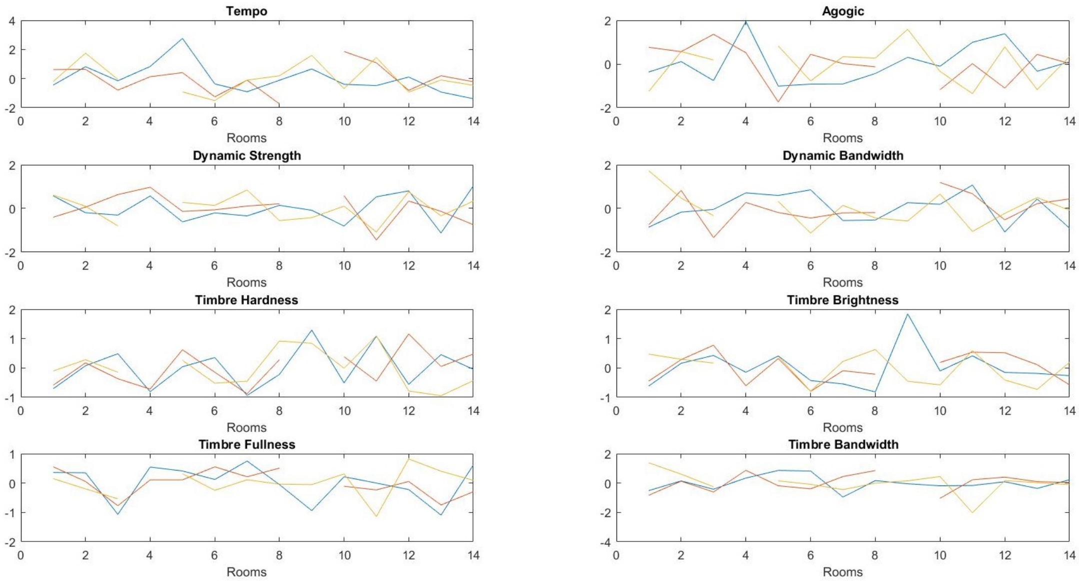


Figure 4.4: Performance characteristics, in blue the first piece for the brass quartet, yellow for the second piece, orange for the string quartet. Some recordings were not calculated part the recording was corrupted (Room 4 for the first piece of the brass quartet, Room 9 for the string quartet)

# Chapter 5

# Statistical analysis

The goal of this chapter is to explain how the relation between the performance characteristics and acoustical coefficients was established in this project, and which statistical models were used for this purpose. In the following, the performance characteristics will be considered as the response variables (output) and the acoustic parameters as explanatory variables (input).

## 5.1 Hypothesis on the data structure

First of all, classical regression models where put aside for several reasons. Firstly, the problem is clearly a multivariate one and a common regression analysis is not designed for this type of problem. [10] [26]. Moreover, in a common regression model it is considered that every observation is independent to another, which is clearly not the case here. It was desirable that the model reflect the nested aspect of the data structure. Several hypothesis on the recordings were indeed made :

- A musician tends to have the same strategy of adaptation whatever the pieces
- Musicians tends to influence each other in the same quartet
- In the same room every musician tends to have a similar strategy of adaptation

This results in a nested structure of this form :

Table 5.1: Structure of the recordings

| Level 4 | Level 3 | Level 2 | Level 1 |
|---|---|---|---|
| Rooms | Quartet 1 | Musicians 1 | Piece 1 |
| | | | Piece 2 |
| | | Musicians 2 | Piece 1 |
| | | | Piece 2 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |

In the following, to reflect this structure the different recordings will be considered in levels :

- The first group level will be for the recordings from the same room, quartet,musician and piece. Constituting the "piece" level.
- The second group level will be for the recordings from the same room, quartet and musician. The "musician" level.
- The third group level will be for the recordings from the same room and quartet. The "quartet" level.
- The fourth group level will be for the recordings from the same room. The "room" level.

The fact that the smaller groups are contained in one bigger is why we called the data structure a nested structure.

## 5.2 Hierarchical Linear Models

### 5.2.1 Theory

Hierarchical Linear Models (HLM) is a multivariate regression model assuming that the data set is hierarchical or more specifically nested, as it is the case here.

It is considered that there is :

- $i \in (1..n_i)$ first-level items ("piece" level)
- in each $j \in (1..n_j)$ second-level groups ("musician" level)
- which are contained in each $k \in (i..n_k)$ third-level groups ("quartet level")
- which are contained in each $l \in (i..n_l)$ fourth-level groups ("room level")

Between the first and second level of the HLM we consider the following regression equation reflecting the grouping effect:

$$Y_{ijkl} = \beta_{0jkl} + \sum_{n=1}^{N_1} \beta_{njkl} X_n + e_{ijkl}$$

Y : the outcome variable here the performance parameter X : the $N_1$ explanatory variables associated to the level

For each item in the first level group there is :

- The constant term (named "intercept") $\beta_{0jkl}$
- The regression coefficients (named "slopes") $\beta_{njkl}$
- The residual error term $e_{ijkl}$

The intercept and slopes vary across the groups of the second level, we consider the following regression equations :

$$\beta_{0jkl} = \gamma_{00kl} + \sum_{n=1}^{N_2} \gamma_{0nkl} Z_n + u_{0jkl}$$
$$\beta_{1jkl} = \gamma_{10kl} + \sum_{n=1}^{N_2} \gamma_{1nkl} Z_n + u_{1jkl}$$
$$\vdots$$
$$\beta_{N_1jkl} = \gamma_{N_10kl} + \sum_{n=1}^{N_2} \gamma_{N_1nkl} Z_n + u_{N_1jkl}$$

- The intercept$\gamma_{0jkl}$
- The slopes $\gamma_{njkl}$
- $u_{ijkl}$ the second-level residual error term

Same regression equations apply between second and third level and the third and fourth level. The different slopes and intercepts are calculated with the maximum likelihood method. To estimate the explanatory power of the HLM the maximum likelihood produce a "deviance" coefficient which reflects the misfit of the model to the data. The lower it is the better the HLM fit the data. Firstly a model with no explanatory variables at each level called "intecept-only" HLM is calculated in order to give a benchmark value to the deviance coefficient.

In our case we consider that the only levels with explanatory variables is the second level (musician) and fourth level (rooms) as the acoustic parameters do not vary only across the rooms but also across the musicians' position. The other regression equations will be considered as "intercept-only" without explanatory variables, however these equations still translate the grouping effect of the data.

The software that was used is *IBM SPSS 24* containing tools for the calculation of this HLM model.

### 5.2.2 Principal Component Analysis

For a first approach, it was decided to limit the number of explanatory variables in order to produce a first perspective on the relation between performance characteristics and acoustical parameters. For this purpose, a Principal Component Analysis (PCA) was ran with the software *IBM SPSS 24* on the acoustical parameters. The principle of a PCA, is to diminish the number of components of a data set by identifying new components not correlated between each other and with the highest variance. The capacity of the new data set to represent the old one is evaluated with the explained variance. Here, we consider the new data set satisfying when the explained variance reach 95 %.

In our case this value is reached for 4 components for each musician.

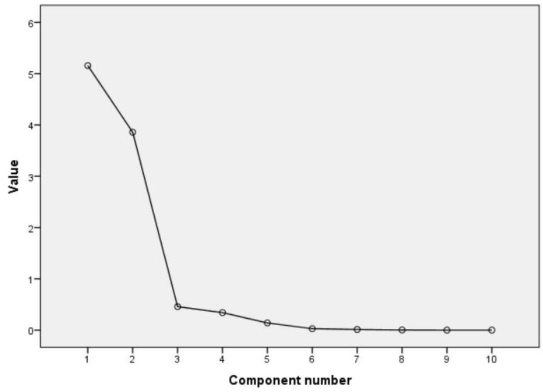


Figure 5.1: Evolution of the explained variance per components of the PCA ($variance = value * 10$) for the acoustic parameters associated to the first trumpet player from *SPSS* the other PCA performed on the quartet produced similar results

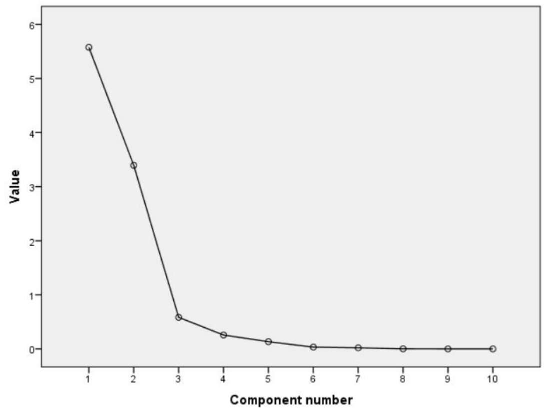


Figure 5.2: Evolution of the explained variance per components of the PCA ($variance = value * 10$) for the acoustic parameters associated to the first violin player from *SPSS* the other PCA performed on the quartet produced similar results

The loadings of the different acoustical parameters is then identified in the different components of the PCA. The parameter with the highest loading is identified in each of the component, these 4 parameters will then be the one used as explanatory variables in the HLM model.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| EDT_VIOLIN1 | -,272 | ,346 | ,860 | -,184 |
| RT_VIOLIN1 | -,060 | ,309 | ,872 | -,337 |
| C80_VIOLIN1 | -,206 | -,890 | -,391 | ,100 |
| G_VIOLIN1 | ,740 | ,664 | ,058 | ,055 |
| ST_early_VIOLIN1 | ,826 | ,449 | -,201 | -,009 |
| ST_late_VIOLIN1 | ,359 | ,870 | ,331 | -,004 |
| G_e_VIOLIN1 | ,907 | ,141 | -,310 | ,201 |
| G_l_VIOLIN1 | ,541 | ,814 | ,207 | -,004 |
| G125_VIOLIN1 | ,803 | ,433 | ,019 | ,351 |
| BR_VIOLIN1 | ,214 | -,047 | -,380 | ,899 |

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| EDT_TRUMPET1 | ,166 | -,916 | ,269 | ,236 |
| RT_TRUMPET1 | ,241 | -,920 | ,010 | ,251 |
| C80_TRUMPET1 | -,860 | ,469 | -,020 | ,191 |
| G_TRUMPET1 | ,956 | ,247 | -,033 | ,107 |
| ST_early_TRUMPET1 | ,838 | ,349 | -,334 | -,020 |
| ST_late_TRUMPET1 | ,924 | -,351 | ,061 | -,128 |
| G_e_TRUMPET1 | ,185 | ,916 | -,026 | ,347 |
| G_l_TRUMPET1 | ,983 | -,142 | ,011 | -,067 |
| G125_TRUMPET1 | ,895 | ,388 | ,066 | ,142 |
| BR_TRUMPET1 | ,244 | ,799 | ,514 | -,111 |

Figure 5.3: Loadings for each of the 4 components of the PCA for the first violin player and first trumpet player

For the different musicians the parameters who were the most often the highest or second highest loadings per components : $ST_{late}$, $G_{early}$, $RT30$, $BassRatio$. Those will be used as explanatory variables in the HLM model.

## 5.3 Use of predictors

Several studies mentioned the possibility of a quadratic relation between reverberation time and tempo. (Kato et al. (2007) : Musicians' adjustments of performance to room acoustics. In : Proc. of the 19th ICA, Madrid). That is why different HLM models with the acoustical parameters used as linear variables or quadratic variables were made. Every $2^4$ possible configuration were calculated for each of the 8 performance characteristics, out of the 128 models the configuration with the highest explanatory power was kept.

Table 5.2: Configuration with the highest explanatory power, HLM used for the results

| | **RT** | **ST late** | **G early** | **Bass Ratio** |
|---|---|---|---|---|
| **Tempo** | Quadratic | Quadratic | Linear | Linear |
| **Agogic** | Quadratic | Quadratic | Linear | Linear |
| **Dynamic strength** | Quadratic | Linear | Linear | Linear |
| **Dynamic bandwidth** | Linear | Linear | Linear | Linear |
| **Timbre (soft - hard)** | Linear | Linear | Quadratic | Quadratic |
| **Timbre (dark - bright)** | Quadratic | Linear | Quadratic | Linear |
| **Timbre (lean - full)** | Linear | Quadratic | Linear | Linear |
| **Timbral bandwidth** | Linear | Linear | Quadratic | Linear |

## 5.4 Intraclass correlation

While calculating the "intercept-only" model, it is possible to evaluate the proportion of variance explained by the structure, and to check if the presence of some level in the structure is relevant. To do so, for each level the Intraclass Correlation Coefficient ($ICC$) is calculated if $ICC < 0.3$ then the level can be suppressed without influencing the explanatory power of the HLM.

$$ICC = \frac{Variance\ of\ the\ residual\ error\ of\ the\ level}{Sum\ of\ the\ variances\ of\ the\ residual\ error\ of\ each\ level}$$

All these variances are calculated via *IBM SPSS 24*.

In our case for each level $ICC$ was over 0.3 confirming that every level had its importance for the explanatory power of the model.

# Chapter 6

# Results

The results provided with *IBM SPSS* enlighten relevant relations on 3 characteristics. The worthiness of a relation is measured in *IBM SPSS* with the p-value a value below 0.05 on a relation is considered relevant.

### Agogic

Confirming previous studies [13] [3] [22], musicians tend to use less agogic in rooms with higher reverberation time. Moreover the presence of early support seems to increase the use of Agogic, as the musicians are more able to hear themselves.

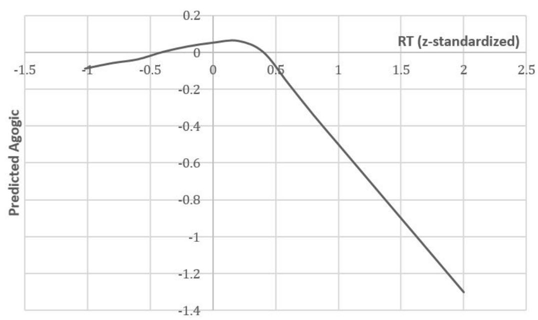


Figure 6.1: Regression curve from *SPSS* of Agogic in function of RT, Regression coefficient with $RT^2 : \beta = -0.28$

### Loudness, Dynamic Strength

Surprisingly no relation was enlighten between support and dynamic strength, the fact of hearing its own instrument "louder" has not influenced the musician in there way of playing. However loudness was lower in rooms with high reverberation time, studies conducted by Kato [13] shown it is a common strategy in order to "hold back" the sound in order to not blurring the different notes.

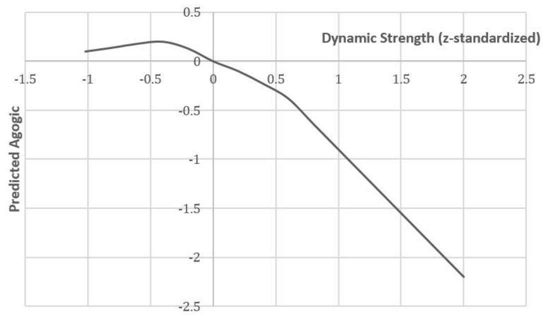


Figure 6.2: Regression curve from *SPSS* of Loudness in function of RT, Regression coefficient with $RT^2$ : $\beta = -0.40$

### Timbre Fullness

Rooms with strong bass ratio lead to performances with a more fuller timbre, the reason of such a result is still unclear.

# Chapter 7

# Conclusion, Perspectives

## 7.1 Summary of the work accomplished

During this internship the main problem was the data set available. It clearly appeared that the number of recordings available was not enough in order to establish relations on how musicians interact between each other in different acoustical configurations. The goal of the project shifted and the focus was put on the method and how it is possible to relate performance characteristics and acoustic parameters. Therefore, the data set was used in order to validate the method more than establish relations.

First of all, it was important to study the stage acoustics theory in order to identify acoustic parameters who are relevant to a musician's point of view, such as support parameters which was rarely used in previous research. Support parameters revised by Wenmaekers [28] had to be integrated to *Raven* as it was not considered as standard parameters (Appendix C).

Once the acoustic parameters were extracted, a method based on the results of the study realised by Weinzierl & Maempel [24] was developed so that the performance of each recordings was evaluated as fully as possible. The first step was to make an automatic onset detection procedure. To do so, a software developed by Lerch [15], in addition to the algorithms converting MIDI to Audio file of *Ableton Live* were employed. It was crucial that the output of the detection required a minimum human intervention for the correction, thereby the method could be applied to a large data set. Once the onset grid of each recordings was calculated the audio features could be extracted and the performance characteristics calculated following the results of the study realised by Weinzierl & Maempel [24].

In the end, it was essential to develop a statistical model that can reflect the structure of the data set as from one recording to the other certain similar conditions could influence the analysis. That is why hierarchical linear models were studied then tested on the recordings available. The few relations that was enlighten by the model correlate to previous studies which seems to show that the method is meaningful and could be applied to larger data set for further research.

## 7.2 Further research

As it was mentioned, a larger data set could led to interesting results. Apart from establishing general results about the performance of the ensemble as a whole, it could be possible to study the influence of the type of instrument and the influence on the interaction between the musicians inside an ensemble.

Furthermore, conducting interviews on the musicians in order to confront their own feelings on the acoustical conditions could also give additional information on the recordings and their integration in the model could enhance the accuracy of the results.

# Appendix A : ISO 3382-1 Recommendations

There are five groups or types of quantities within each group there is often more than one measure, but values of the different quantities in each group are usually found to be strongly correlated with each other. Thus, each group contains a number of approximately equivalent measures and it is not necessary to calculate values of all of them; nevertheless, at least one quantity should be included from each of the five groups.

| Subjective listener aspect | Acoustic quantity | Single number frequency averaging [a] Hz | Just noticeable difference (JND) | Typical range [b] |
|---|---|---|---|---|
| Subjective level of sound | Sound strength, $G$, in decibels | 500 to 1 000 | 1 dB | –2 dB; +10 dB |
| Perceived reverberance | Early decay time (EDT) in seconds | 500 to 1 000 | Rel. 5 % | 1,0 s; 3,0 s |
| Perceived clarity of sound | Clarity, $C_{80}$, in decibels | 500 to 1 000 | 1 dB | –5 dB; +5 dB |
| | Definition, $D_{50}$ | 500 to 1 000 | 0,05 | 0,3; 0,7 |
| | Centre time, $T_S$, in milliseconds | 500 to 1 000 | 10 ms | 60 ms; 260 ms |
| Apparent source width (ASW) | Early lateral energy fraction, $J_{LF}$ or $J_{LFC}$ | 125 to 1 000 | 0,05 | 0,05; 0,35 |
| Listener envelopment (LEV) | Late lateral sound level, $L_J$, in decibels | 125 to 1 000 | Not known | –14 dB; +1 dB |

[a] The single number frequency averaging denotes the arithmetical average for the octave bands, except for $L_J$ which shall be energy averaged [see (A.17)].

[b] Frequency-averaged values in single positions in non-occupied concert and multi-purpose halls up to 25 000 $m^3$.

| Subjective listener aspect | Acoustic quantity | Single number frequency averaging Hz | JND (just noticeable difference) | Typical range |
|---|---|---|---|---|
| Ensemble conditions | Early support, $ST_{Early}$, in decibels | 250 to 2 000 | Not known | –24 dB; –8 dB |
| Perceived reverberance | Late support, $ST_{Late}$, in decibels | 250 to 2 000 | Not known | –24 dB; –10 dB |

That is why, in this project the value used for the parameters will be the average between third octave 500Hz and 1000Hz. Excepting the support parameters averaged between 250Hz and 2000Hz

# Appendix B : Performance attributes, Relation with statistical predictors of audio features

Table 7.1: Vocabulary established by Weinzierl & Maempel

| **Attributes (German)** | **Pole labels** | **Attributes (English)** | **Pole labels** |
|---|---|---|---|
| Klangfarbe | weich-hart | Tone colour | soft-hard |
| Klangfarbe | dunkel-hell | Tone colour | dark-bright |
| Klangfarbe | schlank-voll | Tone colour | lean-full |
| Klangfarbliche Bandbreite | groß-klein | Timbral bandwidth | small-large |
| Phrasierung | kleinteilig-weiträumig | Phrasing: width | narrow-wide |
| Phrasierung | schwach-stark | Phrasing: strength | weak-strong |
| Lautstärke | leise-laut | Loudness | gentle-loud |
| Dynamik | gering-hoch | Long-term dynamics | weak-strong |
| Binnendynamik | gering-hoch | Short-term dynamics | weak-strong |
| Tempo | langsam-schnell | Tempo | slow-fast |
| Agogik | wenig-viel | Agogic | weak-strong |
| Rhythmisierung | unprägnant-prägnant | Rhythmisation | weak -concise |
| Artikulation | gebunden-abgesetzt | Articulation | legato-staccato |
| Artikulatorische Bandbreite | klein-groß | Bandwidth of articulation | small-large |
| Musikalischer Ausdruck | schwach-stark | Musical expression | weak-strong |
| Gesamteindruck | gefällt nicht-gefällt | Overall impression | dislike-like |

Table 7.2: Performance characteristics in function of audio features predictors (Results from Weinzier & Maempel) Some of the characteristics are not present because the explained variance of the relation was considered to low to be relevant

| **Performance characteristics** | **Predictors** | **Coefficients** |
|---|---|---|
| Tempo | Intercept (Constant value) | 4.833 |
| | IOI log mn | -0.651 |
| | IBI log qu10 | -0.309 |
| Agogic | Intercept | -2.005 |
| | IOI_{norm} log qu_{10-90} | 0.441 |
| | TMP_{bar} | 0.305 |
| | SF qdr mdn | 0.215 |
| Dynamic strength | Intercept | -1.048 |
| | dBA qdr mde | -0.451 |
| | SF lin qu10 | -0.45 |
| | ZwiDIN log mn | 0.412 |
| | VUM qdr gmn | 0.199 |
| Dynamic bandwidth | Intercept | 2.403 |
| | ZwiITU lin qu10-90 | 0.612 |
| | ITU1770 log qu10 | -0.67 |
| | dBA lin qu10-90 | -0.446 |
| | MFCC0 qdr qu90 | 0.26 |
| Timbre (soft - hard) | Intercept | 1.455 |
| | SR lin gmn | 1.066 |
| | ZwiITU lin mn | 0.629 |
| | SR log qu10 | -0.458 |
| | ZwiITU lin qu90 | -0.508 |
| Timbre (dark - bright) | Intercept | 0.955 |
| | SR lin qu90 | 0.515 |
| | MFCC1 log mde | -0.23 |
| | MFCC2 qdr mdn | 0.299 |
| Timbre (lean - full) | Intercept | -0.045 |
| | SR lin qu90 | -0.321 |
| | MFCC1 qdr mdn | 0.369 |
| | MFCC3 log mde | 0.36 |
| | TMP_{note} log mde | -0.289 |
| Timbral bandwidth | Intercept | -0.238 |
| | ITU1770 log qu10 | -0.364 |
| | IBI log std | 0.316 |
| | ITU1770 qdr qu10 | -0.284 |
| | ZwiDIN qdr mn | 0.391 |

# Appendix C : Add-on to Raven : function for the extraction of support parameters

```
function [ST_e, ST_l] = getSupport(obj, averageOverReceivers,
    averageOverFrequencies, afterDIN, sourceID)
            if nargin < 2
                averageOverReceivers = 0;
            end
            if nargin < 3
                averageOverFrequencies = 0;
            end
            if nargin < 4
                afterDIN = 0;
            end
            if nargin < 5
                sourceID = 0;
            end

            % calculate schroeder curve
            edc = obj.getSchroederCurve('nodB', 'nonorm', '
                notimecorrect');
            if ~iscell(edc)
                edc = {edc};
            end

            % calculate sound speed
            c = calculateSoundSpeed(obj.getTemperature(), obj
                .getHumidity(), obj.getPressure());



            for iRec = 1 : numel(edc(sourceID+1,:))
```

```
        if ~isempty(edc{sourceID+1,iRec})
            for iFreq = 1 : size(edc{sourceID+1,iRec}, 2)
                % calculate source to receiver distance and arrival time of direct sound
                srcPos = obj.getSourcePosition(sourceID);
                recPos = obj.getReceiverPosition(iRec-1);
                source_receiver_distance = norm(recPos - srcPos);
                directSoundTime = source_receiver_distance / c;

                % find the exact integration limit in the histogram

                integral_end_time_10 = (directSoundTime + 0.010);
                % integral time_20 can be changed to 10ms because less loss
                % when reflective surfaces are closer than 4m
                integral_end_time_20 = (directSoundTime + 0.020);
                integral_end_time_100 = (directSoundTime + 0.100);
                integral_end_time_1000 = (directSoundTime + 1.000);
                last_time_slot_10 = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_end_time_10, 1) - 1;
                last_time_slot_20 = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_end_time_20, 1) - 1;
                last_time_slot_100 = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_end_time_100, 1) - 1;
                last_time_slot_1000 = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_end_time_1000, 1) - 1;


                % relative portion is same at every
```

```
border
% IR resolution of 10ms
relativePortionOfLastTimeSlot_10 = rem(integral_end_time_10, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);
relativePortionOfLastTimeSlot_20 = rem(integral_end_time_20, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);
relativePortionOfLastTimeSlot_100 = rem(integral_end_time_100, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);
relativePortionOfLastTimeSlot_1000 = rem(integral_end_time_1000, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);

% detect energy in last time slot
energyInLastTimeSlot_10 = edc{sourceID+1,iRec}(last_time_slot_10, iFreq) - edc{sourceID+1,iRec}(last_time_slot_10 + 1, iFreq);
energyInLastTimeSlot_20 = edc{sourceID+1,iRec}(last_time_slot_20, iFreq) - edc{sourceID+1,iRec}(last_time_slot_20 + 1, iFreq);
energyInLastTimeSlot_100 = edc{sourceID+1,iRec}(last_time_slot_100, iFreq) - edc{sourceID+1,iRec}(last_time_slot_100 + 1, iFreq);
energyInLastTimeSlot_1000 = edc{sourceID+1,iRec}(last_time_slot_1000, iFreq) - edc{sourceID+1,iRec}(last_time_slot_1000 + 1, iFreq);

% total energy in the impulse response
totalEnergy = edc{sourceID+1,iRec}(1, iFreq);

% energy in the first 50ms/80ms after the direct sound
energy10ms = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_10, iFreq) +
```

```
relativePortionOfLastTimeSlot_10 * energyInLastTimeSlot_10;
                    energy20ms = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_20, iFreq) + relativePortionOfLastTimeSlot_20 * energyInLastTimeSlot_20;
                    energy100ms = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_100, iFreq) + relativePortionOfLastTimeSlot_100 * energyInLastTimeSlot_100;
                    energy1000ms = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_1000, iFreq) + relativePortionOfLastTimeSlot_1000 * energyInLastTimeSlot_1000;

                    % calculate ST
                    ST_e{iRec}(iFreq) = 10*log10(energy100ms - energy20ms);

                    if nargout > 1

                    ST_l{iRec}(iFreq) = 10*log10(energy1000ms - energy100ms);


                    end
                end
            else
                ST_e{iRec} = [];
                if nargout > 1
                    ST_l{iRec} = [];
                end
            end
        end


        if averageOverReceivers
            ST_e = obj.averageOverReceivers(ST_e);
            if nargout > 1
                ST_l = obj.averageOverReceivers(ST_l);
            end
        end

        if averageOverFrequencies
            if nargin < 4
                afterDIN = 0;
```

```
                end
                ST_e = obj.averageAfterDIN(ST_e, afterDIN);
                if nargout > 1
                    ST_l = obj.averageAfterDIN(ST_l, afterDIN);
                end
            end

            if (numel(ST_e) == 1) && iscell(ST_e)
                ST_e = ST_e{1};
                if nargout > 1
                    ST_l = ST_l{1};
                end
            end
        end

        % ------------------------------------------------------------------------


function [ST_ed, ST_ld] = getSupport_d(obj, averageOverReceivers, averageOverFrequencies, afterDIN, sourceID)
            if nargin < 2
                averageOverReceivers = 0;
            end
            if nargin < 3
                averageOverFrequencies = 0;
            end
            if nargin < 4
                afterDIN = 0;
            end%
            if nargin < 5
                sourceID = 0;
            end

            % calculate schroeder curve
            edc = obj.getSchroederCurve('nodB', 'nonorm', 'notimecorrect');
            if ~iscell(edc)
                edc = {edc};
            end

            % calculate sound speed
            c = calculateSoundSpeed(obj.getTemperature(), obj.getHumidity(), obj.getPressure());

            for iRec = 1 : numel(edc(sourceID+1,:))
```

```
        if ~isempty(edc{sourceID+1,iRec})
            for iFreq = 1 : size(edc{sourceID+1,iRec}, 2)
                % calculate source to receiver distance and arrival time of direct sound
                srcPos = obj.getSourcePosition(sourceID);
                recPos = obj.getReceiverPosition(iRec-1);
                source_receiver_distance = norm(recPos - srcPos);
                directSoundTime = source_receiver_distance / c;

                % find the exact integration limit in the histogram
                % %obacht! integralzeiten
                integral_start_time_e = (0.010 + directSoundTime);
                integral_end_time_e = (0.103 - directSoundTime);
                integral_start_time_l = (0.103 - directSoundTime);
                last_time_slot_start_e = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_start_time_e, 1) - 1;
                last_time_slot_end_e = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_end_time_e, 1) - 1;
                last_time_slot_start_l = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_start_time_l, 1) - 1;
                %last_time_slot_end_l = find(obj.histogram{sourceID+1,iRec}.timevector >= integral_end_time_l, 1) - 1;
                relativePortionOfLastTimeSlot_start_e = rem(integral_start_time_e, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);
                relativePortionOfLastTimeSlot_end_e = rem(integral_end_time_e, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);
                relativePortionOfLastTimeSlot_start_l
```

```
= rem(integral_start_time_l, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);
%relativePortionOfLastTimeSlot_end_l = rem(integral_end_time_l, obj.timeSlotLength/1000) / (obj.timeSlotLength/1000);

% detect energy in last time slot
energyInLastTimeSlot_start_e = edc{sourceID+1,iRec}(last_time_slot_start_e, iFreq) - edc{sourceID+1,iRec}(last_time_slot_start_e + 1, iFreq);
energyInLastTimeSlot_end_e = edc{sourceID+1,iRec}(last_time_slot_end_e, iFreq) - edc{sourceID+1,iRec}(last_time_slot_end_e + 1, iFreq);
energyInLastTimeSlot_start_l = edc{sourceID+1,iRec}(last_time_slot_start_l, iFreq) - edc{sourceID+1,iRec}(last_time_slot_start_l + 1, iFreq);
%energyInLastTimeSlot_end_l = edc{sourceID+1,iRec}(last_time_slot_end_l, iFreq) - edc{sourceID+1,iRec}(last_time_slot_end_l + 1, iFreq);

% total energy in the impulse response
totalEnergy = edc{sourceID+1,iRec}(1, iFreq);

% energy in the first 50ms/80ms after the direct sound
energy_start_e = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_start_e, iFreq) + relativePortionOfLastTimeSlot_start_e * energyInLastTimeSlot_start_e;
energy_end_e = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_end_e, iFreq) + relativePortionOfLastTimeSlot_end_e * energyInLastTimeSlot_end_e;
```

```
                    energy_start_l = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_start_l, iFreq) + relativePortionOfLastTimeSlot_start_l * energyInLastTimeSlot_start_l;
                    %energy_end_l = totalEnergy - edc{sourceID+1,iRec}(last_time_slot_end_l, iFreq) + relativePortionOfLastTimeSlot_end_l * energyInLastTimeSlot_end_l;

                    % calculate ST
                    ST_ed{iRec}(iFreq) = 10*log10(energy_end_e - energy_start_e);

                    if nargout > 1

                    ST_ld{iRec}(iFreq) = 10*log10(totalEnergy - energy_start_l);

                    end
                end
            else
                ST_ed{iRec} = [];
                if nargout > 1
                    ST_ld{iRec} = [];
                end
            end
        end


        if averageOverReceivers
            ST_ed = obj.averageOverReceivers(ST_ed);
            if nargout > 1
                ST_ld = obj.averageOverReceivers(ST_ld);
            end
        end

        if averageOverFrequencies
            if nargin < 4
                afterDIN = 0;
            end
            ST_ed = obj.averageAfterDIN(ST_ed, afterDIN);
            if nargout > 1
                ST_ld = obj.averageAfterDIN(ST_ld, afterDIN);
            end
        end

```

```
          if (numel(ST_ed) == 1) && iscell(ST_ed)
              ST_ed = ST_ed{1};
              if nargout > 1
                  ST_ld = ST_ld{1};
              end
          end
      end

      %
```

# Bibliography

[1] L. Aspöck, S. Pelzer, F. Wefers, and M. Vorländer. A real-time auralization plugin for architectural design and education. In *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, pages 156–161, Berlin, Germany, April 2014.

[2] D. Blum. *The art of quartet playing: The Guarneri Quartet in conversation with David Blum.* Cornell University Press, 1987.

[3] S. Bolzinger, O. Warusfel, and E. Kahle. A study of the influence of room acoustics on piano performance. *Le Journal de Physique IV*, 4(C5):620–623, 1994.

[4] E. Clarke and N. Cook. *Empirical musicology: Aims, methods, prospects.* Oxford University Press, 2004.

[5] J. J. Dammerud. *Stage acoustics for symphony orchestras in concert halls.* PhD thesis, University of Bath, 2009.

[6] C. Flesch. *Die Kunst des Violinspiels: Künstlerische Gestaltung und Unterricht: bind 2.* Verlag von Ries & Erler, 1928.

[7] A. Gade. Musicians ideas about room acoustic qualities. *Acoust. Lab., Tech. Univ. Denmark, Copenhagen, Denmark, Tech. Rep*, 31, 1981.

[8] A. C. Gade. Acoustics of the orchestra platform from the musicians' point of view. *Proceedings of Acoustics for Choir and Orchestra, Royal Swedish Academy of Music*, (52), 1986.

[9] I. Galamian and S. Thomas. *Principles of violin playing and teaching.* Courier Corporation, 2013.

[10] J. J. Hox, M. Moerbeek, and R. van de Schoot. *Multilevel analysis: Techniques and applications.* Routledge, 2010.

[11] Q. J. J. Wersuch anweisung die flöte traversière zu spielen. 1752.

[12] K. Kato, K. Ueno, and K. Kawai. Musicians' adjustment of performance to room acoustics, part ii: Acoustical analysis of performed sound signals. 19, 2007.

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

[14] A. Klapuri et al. Musical meter estimation and music transcription. In *Cambridge Music Processing Colloquium*, pages 40–45, 2003.

[15] A. Lerch. Software-based extraction of objective parameters from music performances. 2008.

[16] A. Lerch. *An introduction to audio content analysis: Applications in signal processing and music informatics*. John Wiley & Sons, 2012.

[17] A. Lindau, T. Hohn, and S. Weinzierl. Binaural resynthesis for comparative studies of acoustical environments. In *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.

[18] G. Naylor. A laboratory study of interactions between reverberation, tempo and musical synchronization. *Acta Acustica united with Acustica*, 75(4):256–267, 1992.

[19] S. Pelzer, L. Aspöck, D. Schröder, and M. Vorländer. Integrating real-time room acoustics simulation into a cad modeling software to enhance the architectural design process. *Buildings*, 4(2):113–138, 2014.

[20] J. Sanders. Suitability of new zealand halls for chamber music. *Marshall Day Acoustics Pty Ltd*, 2003.

[21] L. Savioja and U. P. Svensson. Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*, 138(2):708–730, 2015.

[22] Z. Schärer Kalkandjiev and S. Weinzierl. The influence of room acoustics on solo music performance: An experimental study. *Psychomusicology: Music, Mind, and Brain*, 25(3):195, 2015.

[23] E. D. Scheirer. Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103(1):588–601, 1998.

[24] E. D. Scheirer. Extraction and validation of audio features for musical performance analysis. *Alexander Lerch, Hans-Joachim Maempel, Fabian Brinkmann, Stefan Weinzierl*, 2017.

[25] D. Schröder. *Physically based real-time auralization of interactive virtual environments*, volume 11. Logos Verlag Berlin GmbH, 2011.

[26] B. G. Tabachnick, L. S. Fidell, and S. J. Osterlind. Using multivariate statistics. 2001.

[27] R. Wenmaekers, C. Hak, and L. van Luxemburg. On measurements of stage acoustic parameters: time interval limits and various source–receiver distances. *Acta Acustica united with Acustica*, 98(5):776–789, 2012.

[28] R. H. Wenmaekers, L. J. Schmitz, and C. Hak. Early and late support over various distances-rehearsal rooms for wind orchestras. In *Proceedings of Forum Acusticum, Krakow*, 2014.

[29] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media, 2013.