A study on the influence of acoustics on historically informed performance of baroque music

Nolan Eley

CY Cergy Paris Université

ED 405 - Économie, Management, Mathématiques, Physique et Sciences Informatiques (EM2PSI)

Équipe Traitement de l'Information et Systèmes

A study on the influence of acoustics on historically informed performance of baroque music

Présentée par:

Nolan Eley

Pour obtenir le grade de:

Docteur de CY Cergy Paris Université

Specialité: Sciences et Technologies de l'Information et de la Communication (STIC)

Présentée et soutenue publiquement le 12 avril 2023

Devant le jury composé de:

Malte Kob	*Prof., Detmold Hochschule für Musik*	Rapporteur
Mathieu Paquier	*PU, Université de Brest*	Rapporteur
Mitsuko Aramaki	*DR CNRS, Université d'Aix Marseille*	Examinatrice
Théodora Psychoyou	*MCF HDR, Sorbonne Université*	Examinatrice
Catherine Lavandier	*PU, CY Cergy Paris Université*	Directrice
Brian F.G. Katz	*DR CNRS, Sorbonne Université*	Directeur
Marguerite Jossic	*Chargée de recherche, Musée de la musique*	Invitée

Contents

Acknowledgments

Many people played a role in supporting me these last 3 and a half years during my doctoral thesis to whom I owe a great deal of gratitude. I would like to acknowledge the following individuals and groups, among others, for their invaluable support and guidance throughout my doctoral journey.

First, I would like to thank my advisors, Catherine Lavandier and Brian F. G. Katz for their time, support, and trust throughout this thesis.

I would also like to express gratitude to all of the jury members, Malte Kob, Mathieu Paquier, Mitsuko Aramaki, Théodora Psychoyou, and Marguerite Jossic for their thorough review of this manuscript as well as their valuable feedback and insights.

I also must thank those on the "espaces sonores" team, Franck Zagala, Antoine Weber, David Thery, Gonzalo Villegas, Sarabeth Mullins, Martin Lawless, Elliot Canfield-Dafilou, David Poirier-Quinot, Aidan Meacham, and Julien De Muynke. They have all been so incredibly helpful throughout this process both in and out of the lab.

I also owe a great deal of gratitude to the wonderful community of doctoral researchers and post-docs from the Institut Jean le Rond d'Alembert for the rich social life and support they provided throughout this thesis.

I am also deeply thankful to Alexander Lerch, Jude Brereton, Laura Davey, and Benoit Alary for providing helpful academic and scientific advice during this thesis. And lastly, I am thankful for the many fellow researchers I met at conferences and other events, who engaged in thought-provoking conversations and provided valuable feedback on my work.

Abbreviations

ASW	apparent source width
BPM	beats per minute
BRIR	binaural room impulse response
CAVE	cave automated virtual environment
DCT	discrete cosine transform
DFT	discrete Fourier transform
DTW	dynamic time warping
EDT	early decay time
ESS	exponential sine sweep
EVAA	experimental virtual archaeological-acoustics
FDN	feedback delay network
GA	geometrical acoustics
HDRP	high-definition render pipeline
HIP	historically informed performance
HMD	head-mounted display
HOA	higher-order Ambisonics
HRTF	head-related transfer function
ILD	interaural level difference
IOI	inter-onset-interval
IR	impulse response
ITD	interaural time difference
ITU	International Telecommunication Union
JND	just-noticeable difference
LEV	listener envelopment
LTI	linear time-invariant
MFCC	mel-frequency cepstral coefficient
MIDI	musical instrument digital interface
MIR	music information retrieval
OAI	overall acoustic impression
PCA	principal component analysis
PHEND	The Past Has Ears at Notre-Dame

RBF	radial basis function
RIR	room impulse response
RMS	root mean square
RT	reverberation time
SMA	spherical microphone array
SNR	signal-to-noise ratio
SRIR	spatial room impulse response
STFT	short-time fourier transform
SVM	support vector machine
t-SNE	t-dstributed stochastic neighbor embedding
VAE	virtual acoustic environment
VR	virtual reality
VU	voltage unit

Abstract

During a musical performance, musicians constantly monitor and adjust their playing. What they hear inevitably depends on the acoustics of the room in which the performance takes place. This thesis investigates the interaction between musicians and room acoustics. Of principal interest are baroque musical performances with a historically informed interpretation played in historical versus modern spaces. This thesis takes place within the EVAA (Experimental Virtual Archaeological-Acoustics) project, which is dedicated to exploring the acoustics of historical spaces and their function within culture using novel and experimental methods.

An experiment was undertaken in which musicians (flutists, violists, and theorbists) played several pieces in two real spaces (the Salon des Nobles from the Château de Versailles and the amphitheater from the Cité de la Musique) and in their virtual counterparts. The virtual spaces were based on a virtual reality system developed within the framework of the EVAA project, for which a novel auralization architecture was implemented. An analysis of the musicians' experiences revealed that their playing depends somewhat on the acoustics of the room and also revealed the strengths and weaknesses of the virtual reality system.

A significant portion of the manuscript is devoted to the analysis of musical performances resulting from this experiment. Two primary approaches were undertaken: the first uses a somewhat typical framework relying on low-level features, and the other relies on objective measures derived from musicological principles to highlight higher-level features. The first approach revealed some differences in objective measures as a function of the room, but the second approach made it possible to identify in which dimensions of baroque interpretation the performances differed.

Finally, a listening test was carried out to verify that the differences in playing style, revealed by the objective performance measures in the two rooms, were audible and were in agreement with the measures. In this test, musically educated listeners rated recordings within several aesthetic musical parameters. The comparison between the objective performance measures and the listener ratings revealed fairly good agreement among the flute performances, while for the violists, this agreement was inadequate. This difference between the instruments shows that the influence of the room on historically informed performance of the musicians is relatively subtle, and that there is still room for further investigation within this domain.

Resumé

Au cours d'une performance artistique, les musiciens surveillent et ajustent constamment leur jeu. Ce qu'ils entendent dépend inévitablement de l'acoustique de la salle dans laquelle se déroule la représentation. Cette thèse étudie ainsi l'interaction entre les musiciens et l'acoustique de la salle, en s'intéressant principalement à l'interprétation historiquement informée de la musique baroque jouée dans un espace d'époque ou contemporain. Elle s'inscrit dans le cadre du projet EVAA (Expérience Virtuelle en Acoustique Archéologique) qui se consacre à l'exploration de l'acoustique des espaces, d'un point de vue historique et culturel, à l'aide de méthodes nouvelles et expérimentales.

Une expérience a donc été menée dans laquelle des musiciens (flûtistes, violistes et théorbistes) ont interprété plusieurs pièces dans deux espaces réels (Salon des Nobles à Versailles et Amphithéâtre de la Cité de la Musique) et dans leurs homologues virtuels. Les espaces virtuels s'appuient sur un système de réalité virtuelle développé dans le cadre du projet EVAA, mais pour lequel de nouvelles architectures d'auralisation ont été implémentées. L'analyse des ressentis des musiciens a permis de mettre en évidence que leur jeu s'appuie sur le retour sonore de la salle et de révéler les forces et les faiblesses du système de réalité virtuelle.

Une partie importante du manuscrit est consacrée à l'analyse des mesures de performances musicales issues de cette expérience. Deux approches ont été utilisées: l'une qui utilise un cadre classique d'extraction de caractéristiques acoustiques (bas niveau) et l'autre qui s'appuie sur des caractéristiques acoustiques issues de principes musicologiques appliqués à la musique baroque (haut niveau). La première approche a permis de montrer que les mesures acoustiques de performance étaient différentes d'une salle à l'autres, mais la seconde approche a permis d'identifier les dimensions d'interprétation de la musique baroque sur lesquelles les performances des musiciens se sont révélées différentes en fonction des salles.

Enfin, un test d'écoute a été réalisé afin de vérifier que les différences de jeu, révélées par les mesures de performances dans les deux salles, étaient détectées perceptivement par l'oreille humaine et interprétées de manière concordante avec les mesures. Dans ce test, les auditeurs ayant une formation musicale ont évalué des enregistrements sur plusieurs dimensions esthétiques musicales. La comparaison entre les mesures acoustiques et les évaluations perceptives a mis en évidence une bonne concordance d'interprétation pour le jeu des flûtistes, alors que pour les violistes cette concordance est mise en défaut. Cette différence entre les instruments montre que l'influence de la salle sur le jeu historiquement informé des musiciens est subtile et que le champ d'investigation sur ce sujet est encore très ouvert.

Chapter 1

Introduction

During a musical performance, musicians monitor and adjust their playing based on, among other things, aural feedback. This aural feedback is directly linked to the acoustical properties of the room in which the performance is taking place. This is one reason a performance of a composition by the same musician can vary from one concert to another. The impact of room acoustics on the composition, performance, and perception of music has been recognized for centuries (Schiltz, 2003), and recent decades have seen an increase in controlled empirical investigations aimed at understanding these effects more deeply.

In this thesis, a novel virtual acoustic environment (VAE) was used in addition to real performance spaces to study the impact of room acoustics on musicians' performance. Of novel interest to this research is whether the acoustics of a historical space facilitate the performance of music from the same era. More specifically, the baroque era was studied using a performance space from the Château de Versailles. Additionally, musicians specializing in baroque historically informed performance (HIP) were chosen to participate in the study.

This thesis takes place as part of the experimental virtual archaeological-acoustics (EVAA) framework which is a multidisciplinary set of projects investigating the acoustics of heritage spaces using novel and experimental methods. This project specifically, being a partnership with the research center of the Château de Versailles and the Cité de la Musique, is known as the EVAA_Ver project. Additionally, this work was supported by the Paris Seine Graduate School

for Humanities, Creation, Heritage, Investissement d'Avenir ANR-17-EURE-0021 – Foundation for Cultural Heritage Science.

1.1 Motivation

There is no ideal acoustic setting that is optimal for all types of musical performance. A musician's acoustic needs can depend on a number of factors including the instrumentation and the type of music being performed. For example, Kuhl (1954) found a different optimum reverberation time for music from the Classical era (a little more than 1.5 s) than for Romantic era music (2.1 s). Understanding the optimal acoustics for different historical genres is of clear interest to practitioners and listeners of the genre.

The acoustic needs of musicians are manifold. For musicians in an ensemble, being able to hear each other is important so that they are able adjust their level and synchronize their timing. For soloists, it is important to be able to hear themselves well. They also like to hear the acoustics of the hall in order to better form an idea of what the audience hears so that they can adjust their playing accordingly. Beyond these minimum technical needs there is also a desired aesthetic quality of the room acoustics that should facilitate the musician's ability to fully embody the desired expressive characteristics of the music performance.

The HIP movement generally seeks to instill musical performances with principles derived from historical musicology in order to achieve a performance style which is considered to be more appropriate for the music and the period during which it was written. There is generally consensus in the movement about playing on historical instruments (or facsimiles) and adopting stylistic tendencies of the era, however there has not been much discussion about the role of acoustics in HIP (Boren, 2019). One of the main aims of this thesis is to rectify this by providing contextual empirical evidence.

Furthermore, there is still not consensus on the most effective method for investigating the role of room acoustics in musical performance. Both real and simulated acoustics have been used and there are advantages and disadvantages to both (see section 2.3). While performing in real halls provides the most accurate stimuli, the logistics of gaining access to halls and coordinating with musicians can be a huge burden. Additionally, one's acoustic memory becomes less reliable as time passes, making direct acoustic comparisons between real halls quite difficult (Gade, 2010). Of course, using simulated acoustics can be less realistic, but one has much better control over incidental independent variables. Recent research has suggested, however, that the level of authenticity in current state-of-the-art VAEs can be relied on to perform studies of

this kind (Brereton, 2014; Luizard, Brauer, et al., 2019).

Understanding how musicians interact with acoustics is useful for a number of reasons. First, it can help with pedagogical strategies for teaching music students how best to deal with different acoustic settings. While there is some accepted wisdom on this issue, additional empirical data can help improve these strategies. Second, information on the acoustic needs and desires of musicians can assist acousticians and architects when designing new concert halls, for example. Third, it can assist concert promoters in making informed decisions about the appropriate acoustic setting for different musical acts depending on the ensemble and musical genre. Somewhat related, conductors and musicians can make more informed choices about repertoire for a given acoustic space. And lastly, the findings in this study can provide valuable contributions to the field of musicology by providing additional observational evidence pertaining to already-existing theories on historical music performance.

1.2 Objectives

In response to the issues outlined above, a VAE was employed which includes several novel components. The VAE improves on previous ones in several ways: (i) it is based on a calibrated geometrical acoustics (GA) model which allows source and receiver configurations and positions to be easily changed (ii) it radiates the directional characteristics of the instrument in a dynamic way by tracking the musician's movements in real time and adjusting the radiation of the sound in the virtual acoustic space in response (iii) it includes a complementary immersive visualization of the hall that is adaptively rendered to project the correct perspective depending on the location of the user.

One of the objectives of this thesis is to investigate the practicality of this new VAE for these types of studies. This includes examining the impact and effectiveness of these novel components. Of course, the primary objective of this thesis is to understand the impact of room acoustics on the solo performance of historical baroque music. In order to accomplish this, 10 musicians specializing in historically informed baroque performance performed in four settings, two real halls and two virtual simulations of these halls. Both subjective and objective data were analyzed in the pursuit of understanding the impact that these settings had on their playing. Furthermore, a listening test was performed to better understand the perceptual salience of the measured changes in performance.

1.3 Thesis structure

Chapter 2 - Background outlines the topics necessary for global comprehension of the thesis. Theories of music performance are explored in addition to strategies for objectively analyzing them. A foundational background in room acoustics is reviewed to form a basic understanding. This leads to elucidation of more advanced acoustic topics such as virtual acoustics and GA. Lastly, previous experiments investigating the influence of acoustics on music performance are thoroughly reviewed.

Chapter 3 - EVAA auralization system details the design and development of the novel multimodal immersive environment. This includes the process of selecting rooms and measuring their acoustic properties, the design and calibration of GA models, the calibration of the auralization system, and the implementation of the adaptive visual rendering setup. A brief preliminary study with singers is described which was intended to validate the use of the system for experiments studying music performance.

Chapter 4 - Experiment describes the main experiment in the thesis. The experimental design is described in detail, and the results of the subjective questionnaires given to participants are reported. Finally, some problems with encountered with the auralization system are discussed along with feedback from the participants and suggestions for improvements.

Chapter 5 - Music performance analysis outlines the first of two strategies used to analyze the music performances recorded in chapter 4. This strategy is a fairly typical approach where a large number of low-level features are extracted followed by dimensionality reduction and statistical analysis. Broad trends are discussed among groups of musicians followed by analysis of each individual musician's changes in performance style. Finally, the limits to this type of approach are discussed.

Chapter 6 - Baroque analysis framework details the second main approach towards analyzing the recorded performances. This more experimental approach is aimed at identifying features which are salient in a historically informed performance of baroque music. This approach is first verified on a small set of professional recordings representing distinct styles of Bbroque performance then applied to the set of recordings described in chapter 4. Lastly, a small experiment is described which was intended to quantify the influence of the room sound, which was identified as an unwanted influencing factor, on the proposed features.

Chapter 7 - Listening Test summarizes a listening test aimed at understanding the perceptual salience of the performance differences discovered in chapter 6. The motivation for understanding the perception of different musical performances is described, followed by the methods used to perform the test. Finally, the results and their implications for the larger topics of this thesis are discussed.

Chapter 8 - Conclusion revisits the original objectives of the thesis and discusses them in the context of a summary of the findings. Final conclusions are reported as well as recommendations for future work.

1.4 Publications

The following is a list of publications resulting from the research carried out during this thesis:

Eley, N., Lavandier, C., Psychoyou, T., and Katz, B. F. G. (2023). "Listener perception of changes in historically informed performance of solo Baroque music due to room acoustics." In: *Acta Acustica*. (Submitted February 7, 2023)

Eley, N., Psychoyou, T., Lavandier, C., and Katz, B. F. G. (Oct. 2022). "A Custom Feature Set For Analyzing Historically Informed Baroque Performances." In: *Proceedings of the 24th International Congress on Acoustics.* Gyeongju, pp. 1-8.

De Muynke, J., Eley, N., Ferrando, J., and Katz, B. F. G. (July 2022). "Preliminary analysis of vocal ensemble performances in real-time historical auralizations of the Palais des Papes." In: *Proceedings on the 2nd Symposium of The Acoustics of Ancient Theatres.* Verona, pp. 1-4.

Eley, N., Lavandier, C., Psychoyou, T., Jossic, M., and Katz, B. F. G. (Apr. 2022). "Performance analysis of solo baroque music played in a period and modern hall." In: *Proceedings of the 16th french acoustics congress.* Marseille, pp. 1-6.

Eley, N., Mullins, S., Stitt, P., and Katz, B. F. G. (Sept. 2021). "Virtual Notre-Dame: Preliminary results of real-time auralization with choir members." In: *Immersive and 3D Audio: from Architecture to Automotive (I3DA).* Bologna, pp. 1-6.

Chapter 2

Background

This chapter provides a review of the main topics that are essential to this thesis. It presents the interdisciplinary nature of the thesis covering topics of historical musicology and theories of music performance as well as the analysis of musical performance. The fundamentals of room acoustics are covered in addition to the basics of how virtual acoustic simulations are commonly implemented. Furthermore, a summary is provided of previous studies investigating the interaction between acoustics and performance.

2.1 Musical performance

In his seminal theory on motor control, Lashley (1951) often relied on musical performance as an example of the impressive capabilities of humans to perform controlled complex actions. However, beyond a technical feat, musical performance is also an expressive action. Sloboda (2000) described skilled musical performances as consisting of these two major components. The expressive component, he stated, lies in the intentional variations of certain performance parameters which can vary depending on instrument and musical context. It has been long understood that the most compelling performances are ones which deviate (intentionally) from the set of detailed instructions found in the notated score rather than ones which pay strict adherence to them (Seashore, 1938, p. 9).

The chain of musical communication involves not just the performer. Kendall and Carterette (1990) defined a model of musical communication between the composer, performer, and listener, each with shared and unshared implicit and explicit knowledge. As such, any robust analysis of a musical performance should take this into context, recognizing the perceptual experience of a listener can differ from the intentions of the composer and/or performer.

Palmer (1997) suggested a model of music performance consisting of three main parts: interpretation, planning, and movement. The interpretation stage consists of decoding the notated score and seeking to understand the composer's intention. The planning stage is devoted to formulating an approach towards the music guided by stylistic properties and the musician's own intuition and experience. Lastly, the movement stage consists of the fine motor systems which put these intentions into action. Previously, Palmer (1989) had found a strong connection between the expressive intentions of pianists and the perception of the resulting performances, suggesting that musicians' intentions do translate to the audience, at least to some degree.

Other models of music performance describe it as a kind of feedback loop in which the musician is continually updating their intentions based on mainly auditory feedback (Sloboda, 1982; Ueno, Kato, et al., 2010). However, this feedback also consists of other environmental cues (Todd, 1993). The list of all elements which can influence a musical performance may be infinite, however, Lerch (2008, pp. 7–8) has collected an inventory of the more salient factors which include:

- general interpretative rules and stylistic recommendations
- performance plan and expressive strategy
- the performer's personal, social, and cultural background
- the physical abilities of the performer
- preparation, including rehearsal
- auditory, visual, and tactile feedback
- external influences such as humidity and temperature
- internal influences such as emotional state
- audience reaction

These factors are not always distracting or detrimental to the performance. In fact, a certain degree of anxiety or stress has been judged by listeners to improve the quality of the resulting performance (Wilson and Roland, 2002). While this thesis will focus specifically on the effects of the room acoustics on solo musicians' performance, it is beneficial to interpret the results within this wider context.

Much of the research on music performance has focused on notated western music, partly because the existence of detailed scores provides a helpful analysis framework. The discussed models of musical performance, which work best when discussing solo music performances, also exclude certain types of performance scenarios such as jazz improvisation or sight-reading, which have different objectives. This thesis focuses on a subset of notated western classical music, specifically historically informed performances of french baroque music.

2.1.1 Music performance analysis

As previously mentioned, different performances of a single composition can vary even though they are derived from the same underlying musical score. Identifying and quantifying these differences is the domain of the field of music performance analysis. While this could cover a broad range, including music criticism, of interest to this thesis are systematic, signal processing based approaches to music performance analysis.

Carl Seashore was one of the first researchers to study music performance in a systematic way (Seashore, 1938). He examined the connection between psychological attributes of sound: pitch, loudness, time, and timbre to the physical characteristics of a sound a wave: frequency, amplitude, duration, and form. This idea of separating the musical signal under analysis into the broad categories of pitch, loudness, time, and timbre is still in use today (Lerch, 2012). However, not every expressive musical gesture fits neatly into one of these categories.

More sophisticated analysis became possible with the increasing power of computers and the introduction of the musical instrument digital interface (MIDI) standard in 1983. Using MIDI information became a popular way to analyze musical performances, since it provides a wealth of information including the precise note onset and offset times and the velocity (corresponding to intensity) and pitch of each played note (Rothstein, 1995, p. 8). One disadvantage is that performances must take place on MIDI-capable keyboard instruments. (While other MIDI instruments do exist, they can be notably different from their real counterparts, creating a learning curve which could be an obstacle in any research done on them.) Another disadvantage of MIDI is that the performance data is limited to certain quantized parameters which may fail to fully capture the complex modalities of an expressive musical performance.

Other methods of performance analysis seek to extract useful information directly from the acoustic signal using advanced signal processing methods found in the field of music information retrieval (MIR) (Lerch, 2012). These typically have the goal of revealing similar types of information that MIDI provides such as note onsets, pitch, and intensity information. In addition, there are a wealth of so-called low-level features that can also provide much richer information about, for instance, the distribution of spectral energy. One disadvantage is that this methodology can be more prone to errors and often requires some kind of manual verification during various parts of the process. Another disadvantage is that the interpretability of some of these features can be difficult, since they often do not have a direct musical meaning.

Lerch et al. (2020) provides a good overview of the field of music performance analysis. In it, the authors outline some of the primary efforts of the field which include:

- assessing music performance compared to a score, such as in judging a performance
- comparing one performance to another
- building a model of music performance, for example, in order to improve the synthesis of musical performances
- drawing broad conclusions about the nature of performance for musicological purposes, for example, by studying a corpus of performances

The comparison of performances is relevant to this study, since one of the main interests is in observing if performances change significantly due to a room's acoustics. However, beyond simply how different or similar a set of performances is, it is also important to understand precisely how performances change, based on the room acoustics, and to be able to describe this in a musically meaningful way based on the observed data, a current challenge in the field.

2.1.2 Historically informed performance

The historically informed performance (HIP) movement runs contrary to mainstream performance practice. Withing HIP, musicians, in an attempt to convey a more historically-appropriate playing style, deliberately imbue their performance with stylistic tendencies from the era during which the composition was written. These stylistic tendencies are derived from primary sources of the era, such as treatises and performance manuals.

Some of the first examples of anything resembling HIP took place in the 19^{th} century when Mendelssohn revived a number of J.S. Bach pieces, including the St. Matthew Passion in 1829;

although Mendelssohn believed more in reinterpreting this music in stylistic idioms of the day, rather than in preserving the music in its original state (Haynes, 2007, p. 27). Before this, when musicians, particularly in the Romantic period, performed repertoire from an earlier era, "the idea of deliberately changing their performing style to correspond to the music simply did not occur to them" (Haynes, 2007, p. 26).

Two of the most prominent early proponents of HIP were Wanda Landowska and Arnold Dolmetsch who, active in the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries, both took a serious interest in performing on period instruments (Haynes, 2007, pp. 38–40). HIP resembling how it is known and practiced today began to emerge in the 1960s. Initially, the movement was known as "authentic" performance, but this term eventually came to be rejected in favor of "historically informed," partly because the authenticity of any historical performance is impossible to verify and so "authentic" came to be thought of as misleading and erroneous. Donington (1973, p. 17) summarized this well:

> Any ideal of *absolute* authenticity can only be illusory, and perhaps harmful in so far as it has encouraged a rather puritanical and quite unauthentic underplaying of baroque music [...] in some modern performances. But *substantial* authenticity is a realistic aim, capable of bringing improvements such as have already transformed our modern experience of baroque music. For we are of this modern age; and much has changed which could not be changed back even if we so desired.

Aside from the name of the movement, there have also been disagreements about whether, and to what extent, one should even attempt a reasonably authentic performance of historical music. For example, Paul Hindemith, a prominent composer of the 20$^{\text{th}}$ century, stated in a 1950 speech about J.S. Bach: "We can be sure that Bach was thoroughly content with the means of expression at hand in voices and instruments, and if we want to perform his music according to his intentions we ought to restore conditions of performance of that time" (Hindemith, 1952, pp. 16–19). However, the well-known conductor, Bruno Walter, disagreed, stating, "we can no longer be guided by the number of executants that were under Bach's direction in St. Thomas's Church, Leipzig; we must make allowance for the musical and emotional requirements of the work and the acoustic properties of our large concert-halls or churches" (Walter, 1961).

Much effort is put into fulfilling the composer's intentions when preparing a musical performance. However, in much music of the past, certain conventions were so well understood by performers that writing them in the score was deemed unnecessary, so not every musical detail was included (Lawson and Stowell, 2003, p. 2). Also, during the baroque period, the composer was often directly involved in the performance as musician, or conductor, or both

(Rink, 2002, p. 9), diminishing even more the need for additional instructions in the score. HIP is concerned with uncovering these details which, at one point, were implicit but then were gradually forgotten.

Most practitioners of HIP accomplish this by examining primary sources, such as performance manuals and treatises, to understand how music was performed in previous eras. Thankfully, many of these sources have been meticulously scrutinized by musicologists who have collected the information into helpful references (Donington, 1963; Butt, 2002; Lawson and Stowell, 2003, 2012). This study focuses mainly on the baroque period for which there exists a wealth of literature concerning performance practice.

2.1.3 Baroque performance practice

It is impossible to separate baroque performance style from its surrounding context. The approach to style is very much guided by the sound of the instrument, and much of what makes up baroque performance style is due to the instrumental design and configuration. Modern wind instruments, for example, require more effort to start and stop a note, leading to a more legato playing style, whereas baroque wind instruments do not require such effort, lending themselves more readily to a detached and separated playing style, typical of a historical baroque playing style (Haynes, 2007, p. 52).

Donington (1982, p. 165) was aware of this, claiming that "[t]he sound of baroque music can only be recovered on its own instruments in original state, with the techniques and idioms of its original performers; the style is very largely dependent upon the sound." Understanding this, it is not a leap to make the claim that room acoustics should also play an important role in allowing a musician to fully embody a particular performance style. The baroque period also brought a new class of technically proficient musicians. This is a major reason for the rise of an improvisatory and ornamental style of playing, as this was a way to show off such technical skill (Lawson, 2002, pp. 8–9).

It should be noted that the HIP movement is not immune to trends and that various styles, all claiming to be "historically informed" have appeared throughout the course of the last century. However, since around the 1980s, the view on what constitutes a historically informed performance of baroque music has solidified somewhat. An extensive list of stylistic tendencies of the era have been well catalogued in Ponsford (2012), Donington (1982, 1973), Haynes (2007), and Houle (1987) among others. Fabian and Schubert (2009, p. 39) summarized these performance attributes well, stating that baroque performance style consists of

> ... locally nuanced and clearly punctuated articulation, well defined metric groups

> and strongly projected/inflected rhythmic gestures, shallow and selectively used vibrato, and a general revelling in the characteristics of eighteenth-century instruments (e.g., the uneven bow strokes, the variety of tonguing patterns and their effect on tone qualities).

Donington (1982, p. 167) described additional stylistic features of baroque playing style which he deemed as essential, stating, "[t]here are two basic characteristics of baroque sound which, under whatever conditions of performance, it is necessary to achieve: a transparent sonority, and an incisive articulation."

One essential component specifically of french baroque performance style, which was not embraced everywhere is the presence of *notes inégales*. This is a performance practice in which the musician interprets pairs of notes with equal written duration unequally by extending the length of one note and shortening the other, resulting in a rhythmic interpretation resembling a light swing, at least to modern ears (Houle, 1987, p. 86).

2.2 Acoustics

When a constant sound, such as white noise, is played in a room, the energy, contained by the room, will gradually build until it reaches a steady state. Likewise, when the sound is terminated, the energy will gradually decay until the room is silent. This time-to-decay is known as reverberation time (RT) and is more formally defined by the time it takes for a continuous sound in a room to decrease by 60 dB once terminated (Everest and Pohlmann, 2009, pp. 151–153).

Most room acoustic theories rely on the assumption that a room is a linear time-invariant (LTI) system which can be fully characterized by an impulse response (IR). Specifically, an IR only characterizes a room with the specific source-receiver configuration with which it was recorded. Such IRs characterizing rooms are known as room impulse responses (RIRs). The reverberation time can be derived from an RIR, as can many other parameters used to describe and characterize a room's acoustics.

A room impulse response can be considered conceptually to consist of three distinct segments (Savioja and Svensson, 2015). As illustrated in fig. 2.1, first there is the direct sound which has traveled along a free path from the source to the receiver. This is followed by a group of early reflections which are mostly distinct. These are sound waves which have reflected off of objects such as the floor and walls before reaching the receiver, losing energy in the process. Eventually the number of individual reflections grows to become so dense that a diffuse

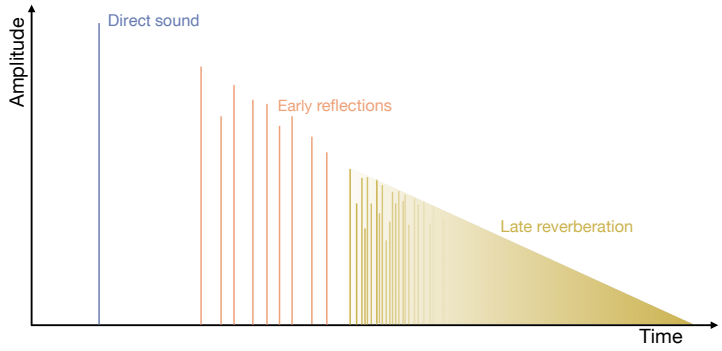


Figure 2.1 Representation of a room impulse response.

reverberation is formed, gradually decaying until all energy has been absorbed by the room's surfaces or air.

One common goal of research in room acoustics has been to improve the sound of concert halls and theaters. This work was pioneered by researchers such as Wallace Clement Sabine (Sabine, 1964) and Leo Beranek (Beranek, 1962). However, much of this work initially focused on the sound perceived from the point of view of the audience. While this is undoubtedly important, research towards understanding and improving the acoustics for performing musicians lagged behind. Marshall et al. (1978) investigated the stage acoustic preferences for musicians, one of the first studies to do so. This has since become an active area of study with many important contributions being made by Gade (1989a,b), who proposed acoustic parameters for evaluating room acoustics from the point of view of musicians (see section 2.2.1.5). This area of research is of particular interest in this thesis, as knowing how the acoustics are perceived by musicians is essential to understanding how it affects their playing.

2.2.1 Room acoustic parameters

Room acoustic parameters are useful indicators of certain attributes of a room's acoustics. These parameters allow for objective comparisons of the acoustics of different rooms. It should be noted that room acoustics are complex and multidimensional, and there is no single acoustic parameter that includes all relevant information. Different parameters may be more or less meaningful depending on the context and requirements. Most of the common room acoustic parameters have been included in the ISO standard 3382-1 (ISO, 2009). These parameters, among others, are well-explained in Gade (2007) and will be summarized below. The just-

noticeable differences (JNDs), which are the generally accepted minimum differences able to be perceived, for each parameter will be reported where available, as well as the best subjective correlate for each parameter.

2.2.1.1 Reverberance

Reverberation time (T) is probably the most apparent acoustic attribute in any concert hall, and it was the first acoustic parameter to be systematically measured and predicted (Sabine, 1964). Adequate reverberation can allow different instruments and voices to mix into an aesthetically pleasing whole, but too much can reduce intelligibility. As stated above, reverberation time (T) is formally defined as the time it takes for a continuous sound, once stopped, to decay by 60 dB. In practice, T is usually calculated based on a smaller portion of the decay curve starting at -5 dB and is then designated accordingly. For example, when T is derived from the portion of the decay curve of -5 dB to -35 dB, it is labeled T_{30}. The best way to calculate T is from the decay rate, $A\frac{\text{dB}}{\text{s}}$, a least squares regression of the relevant portion of the decay curve as

$$T = \frac{60\,\text{dB}}{A\frac{\text{dB}}{\text{s}}} = \frac{60}{A}\text{s}. \tag{2.1}$$

Early decay time (EDT) has proven to be a better descriptor of perceived reverberance than T during running speech or music. It is calculated similarly to T in that it measures the time of a 60 dB decay, but only based on the initial portion of the slope, such as from 0 dB to -10 dB,

$$EDT_{10} = \frac{60}{A_{(0\,\text{dB}\rightarrow -10\,\text{dB})}}. \tag{2.2}$$

Alternatively, EDT can be calculated on a slightly longer portion of the slope, such as from 0 dB to -15 dB in which case it should be delineated as such (e.g. EDT_{15}). The JND for these parameters is about 5%.

2.2.1.2 Clarity

Clarity (C) is the ratio of energies within different portions of the decay curve. It relates to the degree to which a performance is perceived as detailed or blurred, where a higher value would lead to a higher perceived clarity. The two most common time limits are 50 ms and 80 ms and C is labeled accordingly. The 50 ms limit is typically used when speech is the primary use

case, whereas the 80 ms limit is more relevant for music. C_{80} would be calculated as such,

$$C_{80} = 10 \log_{10} \left[\frac{\int_0^{80\,\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{80\,\mathrm{ms}}^{\infty} h^2(t)\mathrm{d}t} \right] \tag{2.3}$$

where $h^2(t)$ is the squared impulse response. The JND for clarity parameters is 1 dB.

2.2.1.3 Sound strength

Strength (G) is a measurement of the effect of the room on the intensity of a signal, and is correlated with the perceived loudness of a signal. It is calculated as the difference between the level of a sound source in a room and the same sound source in a free field recorded at a distance of 10 m:

$$G = 10 \log_{10} \frac{\int_0^{\infty} h^2(t)\mathrm{d}t}{\int_0^{t_{dir}} h^2_{10\,\mathrm{m}}(t)\mathrm{d}t}. \tag{2.4}$$

It also exists as early strength (G_{Early}),

$$G = 10 \log_{10} \frac{\int_0^{80\,\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_0^{\infty} h^2_{10\,\mathrm{m}}(t)\mathrm{d}t} \tag{2.5}$$

and late strength (G_{Late}),

$$G = 10 \log_{10} \frac{\int_{80\,\mathrm{ms}}^{\infty} h^2(t)\mathrm{d}t}{\int_0^{\infty} h^2_{10\,\mathrm{m}}(t)\mathrm{d}t}. \tag{2.6}$$

The JND for G is about 1 dB.

2.2.1.4 Spaciousness

Early lateral energy fraction (LF_{Early}) measures the fraction of sound energy arriving laterally (recorded with a figure-of-eight microphone) compared to the total sound energy in the impulse response (recorded with an omnidirectional microphone), calculated as

$$LF_{Early} = \frac{\int_{t=5\,\mathrm{ms}}^{t=80\,\mathrm{ms}} h_1^2(t)\mathrm{d}t}{\int_{t=0\,\mathrm{ms}}^{t=80\,\mathrm{ms}} h^2(t)\mathrm{d}t} \tag{2.7}$$

where $h_1^2(t)$ is the squared impulse response of the figure-of-eight microphone. As lateral energy increases (in the early part of the decay) so does a listener's perception of apparent source width

(ASW), a subjective acoustic indicator often used in listening tests.

Late lateral energy fraction (LF_{Late}) is the level of late-arriving lateral sound and is calculated thus,

$$LF_{Late} = \frac{\int_{t=80\,\mathrm{ms}}^{t=\infty} h_1^2(t)\mathrm{d}t}{\int_{t=0\,\mathrm{ms}}^{t=t_{dir}} h_{10\,\mathrm{m}}^2(t)\mathrm{d}t} \tag{2.8}$$

where $h_{10\,\mathrm{m}^2}(t)$ is the squared impulse response of an omnidirectional microphone at a distance of 10 m from the source. An increase in late lateral strength is correlated with an increase in listener envelopment (LEV), another subjective acoustic indicator. Because it is primarily energy in the low and mid frequencies that contribute to the sense of spaciousness, the LF measures are typically reported as an average across the four octave bands from 125 Hz to 1000 Hz. The JND of LF_{Early} is about 5%, while for LF_{Late} it is not known.

In addition to lateral energy measures, **inter-aural cross correlation** $IACC$ has been found to be correlated with perceived spaciousness. This measure compares the signal at the left and right ears as usually recorded with an acoustic dummy head or in-ear microphones. It is calculated thus,

$$IACC_{t_1,t_2} = max \left| \frac{\int_{t_1}^{t_2} h_L(t)h_R(t+\tau)\mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} h_L^2(t)\mathrm{d}t \int_{t1}^{t2} h_R^2(t)\mathrm{d}t}} \right| \tag{2.9}$$

where h_L and h_R are the impulse responses recorded at the left and right ear, respectively, t_1 and t_2 represent the time interval of the impulse response during which the correlation is calculated, τ is the interval within which the maximum correlation is searched (usually from -1 ms to 1 ms). The $IACC$ can be calculated either on the early reflections ($t_1 = 0$ s and $t_2 =$ 80 ms) or on the late reverberation ($t_1 = 80$ ms and $t_2 =$ a time longer than the reverberation time). The former metric is typically reported as $IACC_{Early}$ while the latter metric is reported as $IACC_{Late}$. Alternatively, if the entire impulse response is used to calculate the metric then it can be reported as $IACC_{All}$. Values are often reported as $1 - IACC$ so that the value increases with dissimilarity, resulting in a value which correlates positively with the impression of spaciousness.

2.2.1.5 Support

Support parameters (sometimes referred to as stage support) are used to describe how musicians perceive the hall from their position on the stage. **Early support (ST_{Early})** has been found to be correlated with how well musicians in an ensemble can hear each other. It is the ratio of

the first 10 ms of the IR (basically the direct sound and floor reflection) to the total energy of the portion of the IR from 20 ms to 100 ms with a source-receiver distance of 1 m,

$$ST_{Early} = 10 \log_{10} \left(\frac{\int_{20\,\mathrm{ms}}^{100\,\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{10\,\mathrm{ms}} h^2(t)\mathrm{d}t} \right). \tag{2.10}$$

Late support (ST_{Late}) is the ratio of the first 10 ms of the IR to the total energy of the portion of the IR from 100 ms to 1000 ms with a source-receiver distance, again, of 1 m,

$$ST_{Late} = 10 \log_{10} \left(\frac{\int_{100\,\mathrm{ms}}^{1000\,\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{10\,\mathrm{ms}} h^2(t)\mathrm{d}t} \right). \tag{2.11}$$

ST_{Late} is a good indicator of how musicians perceive reverberance in the hall from their position on the stage.

2.2.1.6 Timbral properties

Two parameters are commonly used to describe the spectral balance of a hall. Correlating with the warmth of the hall is the **bass ratio (BR)**,

$$BR = \frac{T_{125\,\mathrm{Hz}} + T_{250\,\mathrm{Hz}}}{T_{500\,\mathrm{Hz}} + T_{1000\,\mathrm{Hz}}}. \tag{2.12}$$

Inversely, the **treble ratio (TR)** can be calculated as such,

$$TR = \frac{T_{2000\,\mathrm{Hz}} + T_{4000\,\mathrm{Hz}}}{T_{500\,\mathrm{Hz}} + T_{1000\,\mathrm{Hz}}}. \tag{2.13}$$

These parameters are convenient, single values that tend to correlate with the perceived "warmth" (in the case of BR) or "brightness" (in the case of TR) of the hall.

2.2.1.7 Spatial parameters

In addition to the above parameters, more recently introduced parameters measuring the directionality of early reflections have shown promise as indicators of musicians' stage acoustic preferences. Research by Domínguez (2008), Dammerud (2009), and Guthrie (2014) found that musicians tend to prefer stronger early reflections from the sides compared to early reflections from the ceiling. Based on this research, Panton et al. (2019) studied two parameters, the **top-sides ratio (TS)** and the **top-horizontal ratio (TH)** and found TH to be a good linear predictor of the subjective indicator of overall acoustic impression (OAI). TH compares the

sound energy from the top to the sound energy from sides of the stage, including the wall behind the stage between a lower limit (t_l) and an upper limit (t_u):

$$TH_{t_l-t_u} = 10\log_{10}\frac{\int_{t_l}^{t_u} h_{top}^2(t)\mathrm{d}t}{\int_{t_l}^{t_u}\left[h_{left}(t)+h_{right}(t)+h_{back}(t)\right]^2\mathrm{d}t}. \tag{2.14}$$

TS is similar to TH but it only includes energy from the left and right walls in the denominator. These parameters must be calculated from measurements using a microphone array, such as a spherical microphone array (SMA) capable of capturing directional energy.

2.2.1.8 Parameter calculation

The most widely accepted approach for measuring the decay curve of an impulse response is through the backward integration method proposed by Schroeder (1965) since this helps even out random fluctuations which can be present in recorded RIRs. This integration is done backwards in time,

$$R(t) = \int_t^\infty h^2(t)\mathrm{d}t = \int_0^\infty h^2(t)\mathrm{d}t - \int_0^t h^2(t)\mathrm{d}t. \tag{2.15}$$

Furthermore, since many of these parameters are frequency-dependent, they are reported in octave bands or 1/3 octave bands by decomposing the RIR into these frequency bands then applying the equations to these filtered signals.

2.2.2 Archaeoacoustics

Archaeoacoustics, sometimes referred to as "heritage acoustics" is an interdisciplinary field devoted to studying the acoustics of historical spaces and their role in culture. Such studies can, for example, lead to new theories of how music was used in specific places or can supplement existing theories with empirical data (Boren, 2019).

In the past few decades, virtual acoustic tools (see section 2.4) have been helpful in recreating the acoustics of historical spaces. One major benefit to virtual acoustic solutions is that they allow for the study of spaces which no longer exist as well as the modification of the configuration of spaces which have undergone significant changes in their history to previous historical states, for example.

Vassilantonopoulos and Mourjopoulos (2001) used virtual acoustic methods to recreate the acoustics of ancient greek buildings which no longer exist. Postma and Katz have conducted a number of studies using virtual methods to study the acoustics of culturally important historic

spaces (Postma, Tallon, et al. (2015), Postma, Poirier-Quinot, et al. (2016), Postma, Dubouilh, et al. (2019), and Katz, Poirier-Quinot, et al. (2019)). Boren et al. (2019) created a real-time acoustic rendering of two different eras of the *Thomaskirche* where J.S. Bach was composer and organist for many years in order to study how musicians might perform Bach's music in the two different configurations. In line with some of these previous studies, this thesis intends to study the role of acoustics of a historical space in musical performances of the era.

2.3 Acoustics—performance interaction

As previously mentioned, room acoustics play an integral role in the creation, interpretation, and perception of music. This link between acoustics and performance practice has been observed and documented for centuries (Schiltz, 2003). Historically, strategies and advice for adapting to different acoustic settings have been documented in performance manuals and musical treatises. However, in recent decades the number of empirical investigations into the matter has been increasing, seeking to find which acoustic qualities affect musicians' performance and to what extent.

While there have been a number of thorough studies examining the effect of acoustics on music performance, it is somewhat difficult to infer universal trends from these studies. This is partly due to the difference in methodology among them. For example, some studies have used virtual acoustic environments (VAEs) (Kato et al., 2008, 2015; Fischinger et al., 2015; Amengual Gari et al., 2019), some have used real acoustic environments (Chiang et al., 2003; Luizard, Steffens, et al., 2020), and others have used a mixed approach (Marshall et al., 1978; Kawai et al., 2013; Schärer Kalkandjiev, 2015; Luizard, Brauer, et al., 2019).

Furthermore, the acoustic parameters under examination in each of these studies can vary. While almost all studies have included reverberation time in their list of examined acoustic parameters, there is a wide variety of other acoustic parameters used. Some of the most common are omnidirectional acoustic parameters such as EDT and clarity. Other parameters, identified by Gade (1989a) as being important to musicians' perception of acoustics, such as stage support are sometimes included (see section 2.2.1 for more details on these parameters). More recent studies have indicated the importance of directionality of early reflections as an indicator of OAI to musicians (Guthrie, 2014) and it has been recommended to include some of these parameters when investigating the interaction between musicians and acoustics (Panton et al., 2019).

Some studies have used ensembles (Marshall et al., 1978; Fischinger et al., 2015; Chiang

et al., 2003) while most have focused on soloists (Bolzinger, Warusfel, et al., 1994; Kato et al., 2008; Schärer Kalkandjiev and Weinzierl, 2015; Amengual Gari et al., 2019; Luizard, Brauer, et al., 2019). Furthermore, the repertoire and performance style have varied widely across these studies. As there is evidence that adaptation to acoustics is at least somewhat dependent on musical content (Schärer Kalkandjiev and Weinzierl, 2015), this is an important factor when comparing studies.

Lastly, many different approaches have been used to analyze the musical performances. Some analysis methods use a large number of mostly low-level features which do not have a direct musical meaning (Schärer Kalkandjiev, 2015; Amengual Gari et al., 2019) while others rely on subjective reports from the musicians themselves (Chiang et al., 2003; Panton et al., 2019). Kato et al. (2008) have used bespoke analysis methods to look at specific features which carry some kind of musical meaning, such as the note-on ratio, which the researchers posited is correlated with the sharpness of staccato. Still other studies have used features which are not derived from the musical signal at all. For example, Henrich et al. (2005) and Luizard and Henrich Bernardoni (2020) have measured the glottal behavior of singers to better understand the connection between physical effort and musical expression.

One of the most common performance parameters measured is tempo, since it is relatively easy to calculate and has a direct musical meaning. Perhaps partly because of this, and because reverberation time is one of the most commonly studied acoustic parameters, correlations between tempo and reverberance are among the most commonly reported trends in these studies. However, measuring tempo is not always straightforward since it can be measured in several different ways, including globally and locally. Furthermore, the statistical methods used to analyze these data can have an impact on the comparability of the results.

Even though reverberation time tends to be one of the most commonly reported independent variables, it is still not evident how changes in reverberation time affect a musical performance. There is a long-held musical intuition that as reverberation time increases, tempo should decrease. For example, Quantz (1752, p. 170), suggested playing more slowly in large rooms compared to playing in small chambers to preserve the intelligibility of the music. This advice, of lowering tempo as reverberation time increases, has been reiterated many times, including in Galamian (1962, pp. 9–10) and Meyer (2009, p. 386). Some studies seem to support this intuition (Bolzinger and Risset, 1992; Fischinger et al., 2015; Kato et al., 2015) or at least partially support it (Amengual Gari et al., 2019). However, this trend is not always clearly borne out in the data. Schärer Kalkandjiev and Weinzierl (2015), for example, showed that this relationship only held true to a certain extent. Firstly, this effect was only observed for

compositions which were generally slow, and was not observed in fast compositions. Secondly, this trend was only linear within a certain range, and slower tempos were also used in extremely short reverberation times, perhaps to compensate for the lack of acoustical decay. This non-linear relationship between tempo and reverberation time was also found in Kato et al. (2015). Lastly, there are some studies that found little to no correlation between reverberation time and tempo (Bolzinger, Warusfel, et al., 1994; Luizard, Brauer, et al., 2019).

One important conclusion found in several of these studies is that adaptation strategies can be dependent on musical content and on the individual musician who may exhibit consistent but unique approaches to different acoustic environments (Amengual Gari et al., 2019; Kato et al., 2015; Luizard, Brauer, et al., 2019; Kob et al., 2020).

2.3.1 Review of empirical studies

One of the first studies to examine the interaction between room acoustics and musicians in a systematic way was by Marshall et al. (1978). In this study, researchers manipulated the timing, intensity, and timbre of early reflections to find which reflections were preferred for playing in a small ensemble. That study focused primarily on the acoustic preferences of musicians. It was found that musicians were able to hear subtle differences in early reflections, and that acoustical preferences can be multidimensional. In other words, different acoustic parameters may take precedence depending on the nature of the performance.

Bolzinger and Risset (1992) studied five solo pianists playing in four different real acoustic conditions. The researchers found that, with increasing reverberation time, the musicians tended to play softer, a little slower, and with less use of the sustain pedal. However, a follow-up study was done shortly after by another team using very similar methodology which found no connection between reverberation time and tempo while the correlation between intensity and reverberation time remained (Bolzinger, Warusfel, et al., 1994).

Chiang et al. (2003) evaluated the subjective experience of solo and chamber instrumentalists playing in five concert halls to better understand musicians' perception of stage acoustics. While no objective parameters were extracted from the performances, the study still revealed a notable preference for strong early reflections compared to those previously reported by orchestral musicians. Furthermore, it was found that soloists were more sensitive to certain changes in acoustics compared to ensemble musicians. This strongly supports earlier studies which suggest that the optimal acoustics depend on the specific use case.

The team of Kato, Ueno, & Kawai have performed a series of in-depth studies investigating the effect of room acoustics on musicians' performance (Kato et al., 2008, 2015; Kawai et al.,

2013; Ueno, Kato, et al., 2010). Those investigations utilized an immersive virtual acoustic setup based on measured spatial room impulse responses (SRIRs). The researchers performed sophisticated analysis on the recorded musical performances paired with subjective reports from musicians. The findings of those studies have been somewhat mixed, but indicate that reverberation time had the strongest effect on musicians' performance. In one study (Kato et al., 2015) they found that the tempo was reduced for short and long reverberation times. There was a positive correlation found between the two performance parameters—staccato and vibrato intensity—and the reverberation time. In other studies, reverberation time was also found to have a negative correlation with pedal usage on the piano (Kawai et al., 2013) and a positive correlation with the note-off duration (Kato et al., 2008).

Brereton (2014) explored the interaction of acoustics and singers within a VAE. The main performance parameters under investigation were tempo, intonation, and vibrato. Despite the singers reporting specific adaptation strategies with high confidence, the data showed fairly minor differences. One of the most prominent performance changes measured was in tempo. Evidence of individual adaptation strategies was also found.

Schärer Kalkandjiev (2015) performed a study in two major parts, one in which a soloist played in seven concert halls and another in which 12 solo musicians played in a VAE under 14 different acoustic conditions. A number of mostly omnidirectional acoustical parameters were taken into consideration and some objective music performance analysis was carried out. The results of that study indicated that early decay time (EDT), reverberation time, and late sound strength (an alternative to ST_{Late} proposed by Dammerud, 2009) had the most significant effects on musicians' performance while clarity and ST_{Late} were of little importance. Furthermore, the study showed that different musicians exhibited different strategies for adapting to changes in acoustics. The study confirmed that the reverberation time had a significant effect on the tempo but that this effect was more evident for slow pieces than for fast pieces.

Fischinger et al. (2015) studied a choir performing in three VAEs. The ensemble in that study responded to longer reverberation time by slowing down, aligning with mainstream musical intuition. However, the ensemble performed under the direction of a conductor (who was also hearing the VAE) whose influence on certain performance parameters, notably, tempo is significant. A major achievement in this study was the success of developing an auralization system for an ensemble. Each member of the ensemble was equipped with half-open headphones and a microphone so they could hear themselves and each other as well as the simulated acoustics, and the participants reported that this acoustic setup felt quite natural.

Amengual Gari (2017) studied 11 solo trumpet players performing in a VAE based on

SRIRs. The main acoustic parameters under investigation were T_{30}, EDT, and a modified version of G. The performances were analyzed by extracting 44 low-level features, applying dimensionality reduction, and performing correlation analysis. A subset of players were found to follow the trend of slowing their tempo as reverberation time increased. Beyond that, some consistent but individual adaptations to tempo and dynamics were also observed, but broad general trends were mostly absent.

Luizard led several studies (Luizard, Brauer, et al., 2019; Luizard, Steffens, et al., 2020; Luizard and Henrich Bernardoni, 2020) investigating the role of acoustics on singing. The first study (Luizard, Brauer, et al., 2019) used both real acoustics and simulated versions of these acoustics and found that the adaptations to acoustics were similar among the real halls and the virtual halls, supporting the suitability of auralizations for these types of studies. The overall trends in this study were not significant, however, some individual trends were quite strong, suggesting again that individuals have consistent but unique strategies for adapting to acoustic changes. The second study (Luizard, Steffens, et al., 2020) confirmed this, finding no significant common patterns of adaptation but revealing some significant relations between certain individuals' performances and acoustics.

2.4 Virtual Acoustics

The primary objective for most virtual acoustic systems is to recreate an auditory environment such that it is essentially indistinguishable from reality. The advantage of using such a system for studying the effect of acoustics on musical performance is obvious, as it allows researchers to bypass many of the challenges inherent in field studies. Virtual environments give researchers finer control of the many independent variables at play, by reducing the time between successive stimuli and removing barriers caused by logistics, such as gaining access to concert halls and scheduling multiple sessions with various musicians (Gade, 2010).

2.4.1 Auralization

Auralization can be thought of as analogous to visualization; it is the process of rendering sound fields audible (Kleiner et al., 1993). While there have been several approaches to auralization historically, including the use of scale models, currently, computational approaches are the most common.

There are two main approaches to computing an RIR. The first is a numerical wave-based approach, such as finite-element method or finite-difference time-domain method (Pietrzyk and

Kleiner, 1997; Hamilton and Bilbao, 2018). While these approaches can produce accurate results, they are limited by certain aspects, and tend to be computationally expensive. Therefore, geometrical acoustics (GA) based approaches, which are computationally less expensive, are often used. In GA approaches, sound is modeled to propagate as rays while approximations are made to model certain wave phenomena such as scattering and diffraction. GA approaches also fail to replicate modal behavior which can be an issue for small spaces. Two common GA approaches are ray-tracing (Krokstad et al., 1968) and the image-source method (Allen and Berkley, 1979). In short, wave-based simulations more accurately model the physical behavior of sound waves in a room, but this comes at a high computational cost. By contrast, ray-tracing methods often fail to accurately reproduce certain behaviors of waves, such as diffraction and modal behavior, employing approximations with a moderate computational cost.

2.4.2 Sound field rendering

Sound fields are typically reproduced through headphones or loudspeakers. Binaural rendering is a method to render sound fields over headphones. In binaural reproduction, the left and right channels are processed in a way that reproduces natural localization cues, such as interaural level differences (ILDs), interaural time differences (ITDs), and spectral cues caused by complex interactions of sound waves with the pinnae (Blauert, 1997). One straightforward way to reproduce an auditory scene is to record it with an acoustic dummy head or in-ear microphones, then play it back through headphones. Another method is to convolve an SRIR with a head-related transfer function (HRTF) which is a set of angle-dependent transfer functions that characterize the interaction of sound with the head, torso, and pinnae, resulting in a binaural room impulse response (BRIR). However, since every individual has different anthropometric characteristics, a generic HRTF measured on a dummy head may not properly reproduce the localization cues accurately for an arbitrary listener. Ideally, an individualized HRTF should be used along with proper headphone equalization to ensure the best spatial audio reproduction via headphones. A more immersive rendering should also include head orientation tracking so that a stationary sound source appears to remain stationary as a listener moves their head.

Another common reproduction technique is through a multichannel loudspeaker array. One advantage to this method is that it does not require any HRTF individualization or head-tracking since the directional properties are linked to the actual speaker positions in the reproduction room. One of the most common implementations for reproducing a sound field over a loudspeaker array is through Ambisonics. Ambisonics is a recording, transmission, and reproduction format based on the decomposition of the sound field into spherical harmonics

(Gerzon, 1973). As the order of spherical harmonics increases, the reproduction of the sound field improves, and the minimum number of loudspeakers needed to reproduce the sound field increases. Higher order Ambisonics are still typically less spatially precise than binaural rendering, however. Another important consideration when rendering Ambisonics over loudspeakers are the acoustics of the playback room and the size of the correctly rendered sound field (commonly referred to as the "sweet spot"). Ideally, the acoustics should be as dry as possible so as not to interfere with the signal from the loudspeakers. A major benefit to using Ambisonics is that it can be decoded to an arbitrary loudspeaker setup, although the quality can be degraded for irregular configurations.

2.5 Summary

This chapter reviewed the fundamental theories and models of music performance including the principles of historical baroque performance style which is the focus of this study. Room acoustics and the parameters typically used to quantify and describe them in addition to their subjective counterparts were also reviewed. The relationship between these two subjects was also reviewed, including a survey of empirical studies devoted to examining this relationship. The broad findings from these studies are that the effect of room acoustics on performance is rather subtle, and somewhat dependent on the character of the piece, the musician, and instrument. Additionally, reverberation time was the acoustic parameter with the most consistently measured effect on performance while tempo was the performance parameter most commonly affected. However, this may be partly due to the commonness with which these parameters are included in such studies. Methods for virtual acoustics and their usage in studies examining the effect of acoustics on performance style were also briefly discussed.

In the following chapter (chapter 3), the implementation, refinement, and initial testing of the virtual acoustic system designed for this study will be described. Additionally, chapter 4 details the primary experiment in this study designed to examine the impact of room acoustics on the historically-informed performance of baroque music using both real and virtual acoustics.

Chapter 3

Auralization system

The experimental virtual archaeological-acoustics (EVAA) project is an ongoing research effort devoted to acoustic heritage for which a real-time auralization architecture has been developed. This auralization system was designed to allow people to interact with acoustic spaces in a realistic way, providing a framework which can be used to examine acoustic heritage questions from the point of view of both listeners and performers in a consistent manner. The framework was originally introduced in Katz, Leconte, et al. (2019) and has been expanded upon since. The general system architecture is described in section 3.1. The rest of the chapter describes the implementation and refinement of the auralization system specifically for simulating the two spaces used throughout this thesis, the Salon des Nobles from the Château de Versailles and the amphitheater from the Cité de la Musique (see section 3.2). The acoustics of these two spaces served as the primary independent variables in this thesis which investigated the role of room acoustics in historically informed performance of baroque music. Lastly, a preliminary study is described in detail which was performed in order to test the validity of the system for use with musicians.

3.1 System architecture

The EVAA system is intended to be somewhat flexible in its implementation. For example, it can utilize different orders of Ambisonics and reproduce the sound field through headphones or a loudspeaker array. The following text describes the specific implementation used for this thesis.

The primary objective of the auralization system in this study is to allow a musician to be placed in different acoustic environments in order to investigate the influence of the acoustics on the playing style of the musician. The system was designed to incorporate the instrument's directivity and even adapt to musicians' movements, dynamically changing the orientation of the source directivity in accordance with the orientation of the musician in real time, a feature that is referred to as "dynamic directivity". The processing takes place within a VST3 plug-in which will be referred to as the EVAA plug-in. This plug-in was used as part of a Max/MSP patch which provides input to the plug-in and takes the plug-in's output and routes it to the output device. A schematic diagram of the system can be seen in fig. 3.1.

The system is fed by a monophonic signal of the instrument, recorded by a small, omnidirectional wireless microphone (DPA 4060) which was affixed to the instrument. At least one previous study (Laird, Murphy, et al., 2011) found that using a microphone placed in a fixed position near the instrument yielded an inconsistent capture of the sound which led to discomfort of the musician who then felt constrained while playing. Therefore, a wireless microphone affixed to the instrument was deemed essential. A RØDELink wireless system was used including a TX-BELT transmitter and RX-DESK receiver.

The system creates an auralization from 3rd-order Ambisonic room impulse responses (RIRs) simulated with a calibrated geometrical acoustics (GA) model using a higher-order Ambisonics (HOA) receiver and a coincident source. These GA models are calibrated according to a procedure outlined in Postma and Katz (2015) which is detailed further in section 3.4. Since the direct sound is provided directly by the musician, it is removed from the RIRs beforehand. The floor reflection was left in however, since the floor of the studio is a somewhat absorptive carpet material which would likely not provide a prominent reflection. The direct sound was removed from the RIR by editing the audio files to remove everything up to the first reflection as this would help minimize the latency of the system. These multichannel RIRs are decoded for playback through a 32-channel loudspeaker system using the Spat[1] decoder set to All-Round Ambisonic Decoding (`AllRAD`) and `max` r_E weighting (Zotter and Frank, 2012). The system

[1] https://forum.ircam.fr/projects/detail/spat/

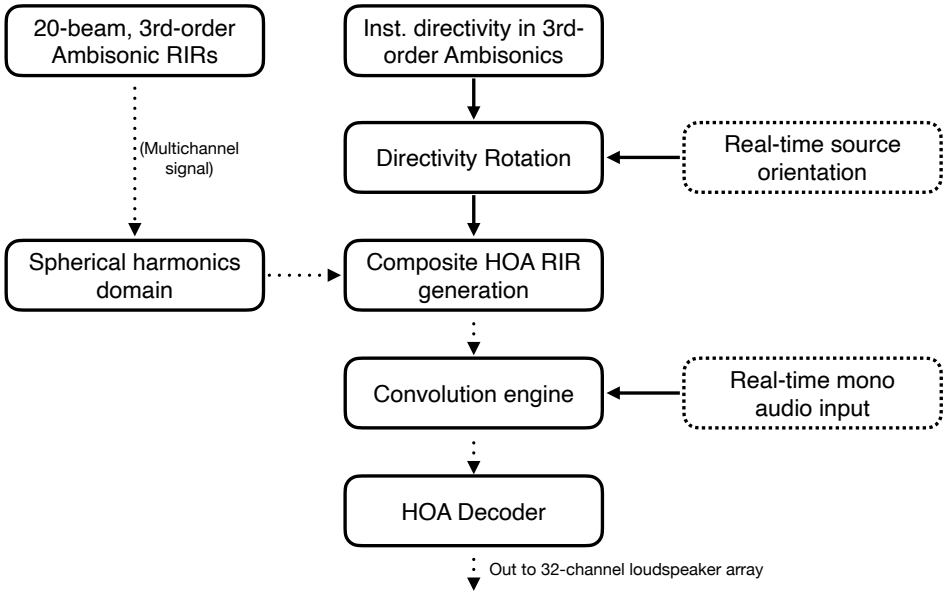


Figure 3.1 A schematic diagram of the EVAA system.

also has the capacity to output the binaural sound through headphones with head-tracking, also using the Spat decoder.

The 32 loudspeakers (ELAC 301) are attached to a cubic frame in three height layers. Eight loudspeakers were situated at floor level and 12 loudspeakers each were situated at the heights of 1.4 m and 2.5 m. It is constructed in the motion capture/virtual reality room at Sorbonne University. The dimensions of the room are 4.09 x 4.48 x 3.07 meters. The room was acoustically dampened and had a T_{30} of 0.14 s at 1 kHz.

3.1.1 Dynamic directivity implementation

3.1.1.1 Source directivity in virtual acoustic environments (VAEs)

It has been shown that the inclusion of directional characteristics of a sound source yields more convincing and plausible auralizations. Wang and Vigeant (2004) showed a difference in perceived reverberation time and clarity as a function of source directivity. Otondo and Rindel (2004) demonstrated that source directivity affected perceived loudness and reverberance. Vigeant et al. (2007) found that the perception of realism increased along with the spatial resolution of the source directivity in auralizations. Furthermore, Kearney (2010) claimed that "for any virtual acoustic recording to be convincing, the directional properties of the source

audio must be considered." It should be noted that the above studies all considered directivity from the point of view of a listener situated in the audience, whereas in the EVAA auralization system, the listener is also the performer. There is very little research on the effect of directivity in auralizations with this type of configuration.

While most VAEs now consider the directional characteristics of the source, a real-time implementation of a dynamic directivity is not common. Postma and Katz (2016) found that auralizations using dynamic directivity were judged as more plausible than those using static directivity from the perspective of a listener in the audience with a source on stage. Arend et al. (2019) developed a VAE which incorporated dynamic directivity but did not study the perceptual importance of the inclusion of such a feature. Regardless of the lack of literature on the perceptual effect of including dynamic directivity for performers, in the pursuit of realism this element was included in the EVAA auralization system. Dynamic directivity, as implemented in this system, follows the orientation of the musician and adjusts the directivity in the virtual acoustics in accordance with these changes. The directivity in this system does not change dynamically as a result of the note played or the phoneme sung as it can in reality (Shabtai, Behler, et al., 2017).

3.1.1.2 EVAA implementation

The source in the GA model is represented by 20 directional sources, or beams, arranged in an icosahedron, each with an outward orientation and directivity such that the sum of all beams yields an omnidirectional directivity with a variation of $\pm$ 0.2 dB. This approach was first introduced in Postma and Katz (2016), then later perceptually verified from an audience listener's perspective in Postma, Demontis, et al. (2017) using 12 beams arranged in a dodecahedron. This was later expanded to 20 beams to produce a higher spatial accuracy. Prior to the study described in section 3.8.2, the system had not been tested from a performer's point of view.

The directivity patterns utilized were derived from Shabtai, Behler, et al. (2017), a database of radiation patterns of 41 musical instruments. The directivity patterns for each of the instruments were captured using a spherical array of 32-microphones, with a radius of 2.1 m. Each instrument was recorded playing a chromatic scale across the entire range of the instrument at two extreme dynamics, very soft (pianissimo, or ***pp***) and very loud (fortissimo, or ***ff***). The directivity patterns for each note at these two dynamic levels were provided as well as averaged directivity patterns in third-octave bands. The directivity pattern can depend on the note played, especially for woodwinds where the sound is radiated through open finger holes or for singers where the directivity is strongly dependent on the phoneme sung. Because of this fact,

this third-octave band representation is imperfect but compact, convenient, and allows for easy use in auralizations. This representation is available as 4^{th}-order spherical harmonics. Acoustic source centering was also applied to each instrument recorded in the database following the procedures in Shabtai and Vorländer (2015) and Ben Hagai et al. (2011).

These instrument directivity patterns and the beam directivities can be converted to spherical harmonics, allowing for smoother processing in the convolution stage and simplified application of dynamic directivity based on orientation changes of the performer via a rotation matrix. An overlap-add uniform-partitioned frequency domain convolution is used as it is a more efficient implementation for real-time usage. The third-octave band representation of the directivities recorded at the ***ff*** dynamic were used. The EVAA plug-in implements the frequency-dependent directivity in octave bands so the third-octave band representations were summed into octave bands prior to implementation such that the power of each octave band equals the total power of its constituent third-octave bands.

The performer's orientation is acquired using an OptiTrack motion capture system and the accompanying Motive (v1.8.0) software. The system utilizes ten cameras (OptiTrack Flex 3) which are positioned throughout the room with redundancy in mind such that at any obstruction will not seriously impede the motion capture system. A motion capture device with markers on it is placed either on the musician's head (in the case of a singer, for example) or on their instrument to track their orientation. Before each session, the motion capture device must be calibrated to define the default starting orientation. This orientation information is then transmitted to the Max patch where it is accessed by the EVAA plug-in for rendering. The constantly changing orientation requires constantly updating RIRs, which are held in separate convolution engines, and in order to avoid discontinuities, are transitioned via cross-fading.

3.1.1.3 Instrument selection

The instruments used in the various studies throughout this thesis include the voice (tenor and baritone), the baroque transverse flute, the viol da gamba, and the theorbo. Most of these instruments do not have perfect matches in the database of radiation patterns. In these cases, a closest match was used. The instruments, and associated details included in the directivity database (Shabtai, Behler, et al., 2017) are discussed below. For all the voices, the soprano directivity was used. For the viol da gamba, a historical cello was used. The cello, made in 2007, was modeled after an instrument made in 1730 and used catgut for the D and A strings and silver-wounded catgut for the C and G strings. The closest match for the theorbo was a concert acoustic guitar. This instrument was made in 1985 and used nylon strings for the

Figure 3.2 A visual model of a top and side view of the Salon des Nobles from the Château de Versailles created by a team of students of the *licence professionelle*, "Patrimoine, Visualisation et Modélisation 3D" at CY Cergy Paris University.

three lowest pitched-strings and roundwound strings for the three highest-pitched strings. The database did include a directivity pattern for a baroque transverse flute, so this was used as the pattern for the flutists in this study. It should be noted though that the precise spacing of the holes on a baroque transverse flute could differ from instrument to instrument and that this could affect the instrument's directivity pattern.

3.2 Room selection

As one of the central questions of this thesis is whether or not a historically appropriate room facilitates historically informed performance, it was necessary to include a room in which musical performances took place during the baroque period. Since this study was limited to solo instrumentalists, the chosen room would ideally be suited for soloists or small ensembles. Additionally, a modern counterpart, suitable for soloists or small ensembles was desired to provide contrast.

The research center at the Château de Versailles was a partner on this project and provided

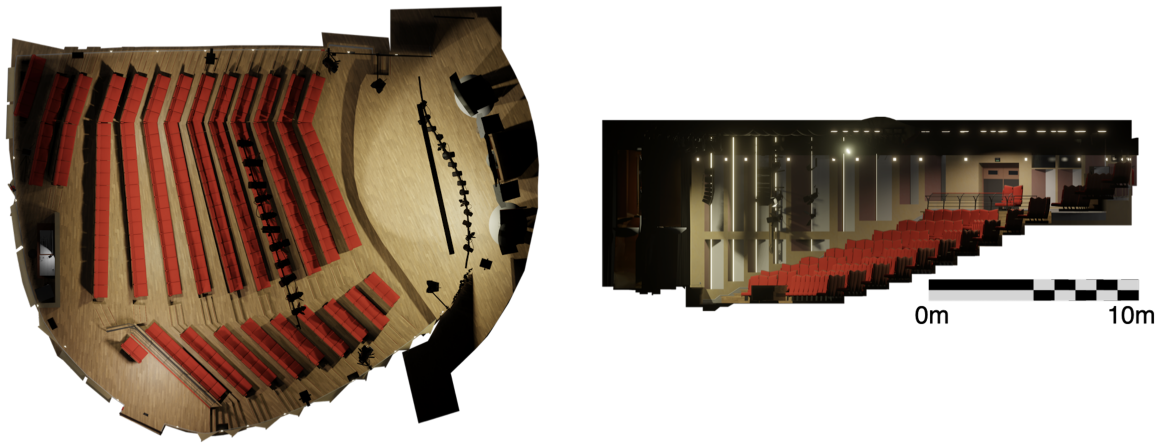
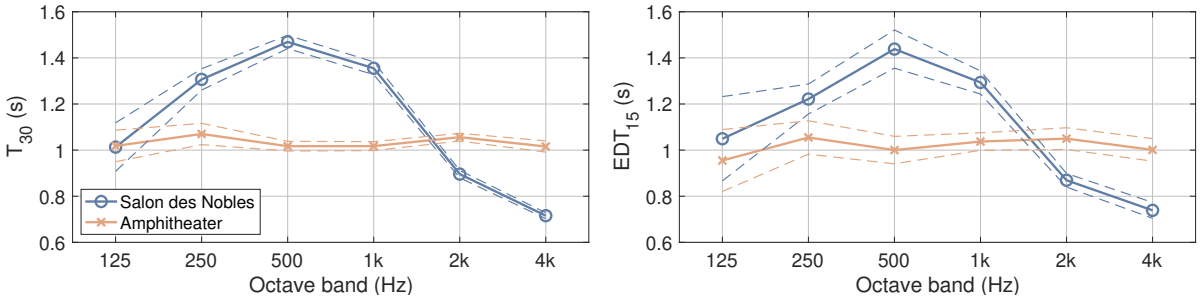


Figure 3.3 A visual model of a top and side view the amphitheater from Cité de la Musique created by Yoann Siohan.

access to the château during the project. Music was an integral part of life at the Château de Versailles, accompanying many of its activities including church services, official ceremonies, walks on the grounds, children's education, bedtime rituals, and nights of entertainment (Baumont, 2007). Marie Leszczyńska, the wife of Louis XV, was an amateur musician and a lover of music. As such, she hosted many concerts in her suite of apartments. One of these rooms, the Salon de la Paix, was often used as a game room in which the queen and her court would indulge in gambling (another hobby of which she was particularly fond), often to the accompaniment of music provided by small ensembles. The Salon des Nobles was another room in which the queen would host concerts regularly which became known simply as the "queen's concerts". While the repertoire for these concerts was not officially recorded, a few examples remain in other primary sources such as diaries, indicating that these concerts typically showcased small ensembles and soloists (Baumont, 2007).

While the music usage documented in both of these rooms would justify using either for this study, the Salon de la Paix is directly adjacent to the Hall of Mirrors, creating a coupled volume with a large and reverberant space. In order to avoid this problematic acoustic configuration,



(a) T_{30} (b) EDT_{15}

Figure 3.4 The (a) measured reverberation time and (b) early decay time, averaged across all source and receiver positions, of the rooms used in this study. The dotted lines represent the standard deviation for all source and receiver positions.

it was decided to use the Salon des Nobles. It is difficult to estimate the seating capacity in the Salon des nobles since the room was not only used for concerts and therefore did not have a permanent seating configuration. However, since the queen would host concerts in this room for her court, it is not likely that a concert would be attended there by more than a few dozen.

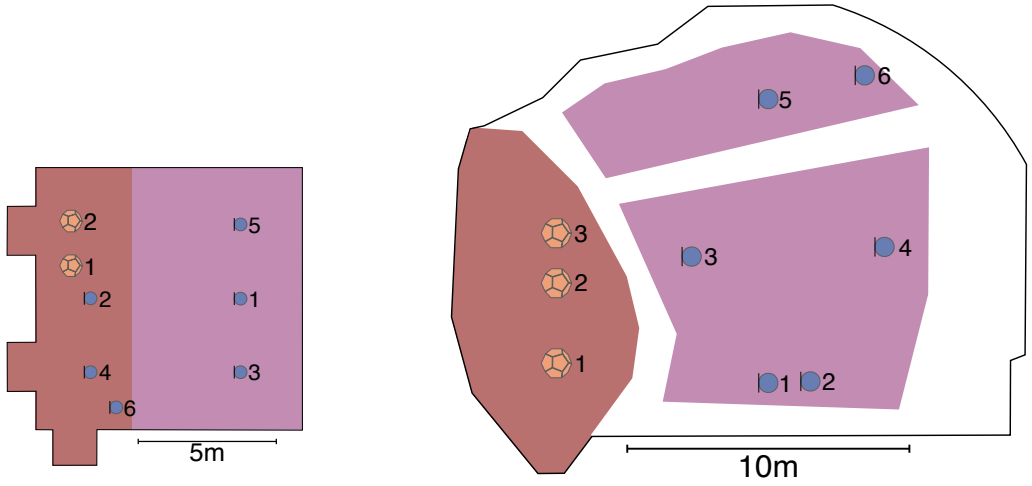
The modern hall that was chosen to provide contrast was the amphitheater from the Cité de la Musique, a group of music-related institutions in the 19th arrondissement of Paris. This small hall, with a seating capacity of approximately 250 is well-suited for solo and chamber music. Digital models of an overhead and side view of both rooms can be seen in figs. 3.2 and 3.3.

3.3 Acoustic measurements

Acoustic measurements were taken of both of the rooms using the exponential sine sweep (ESS) method (Farina, 2007). Because these measurements were intended to provide the necessary data to create and calibrate GA models, only omnidirectional sound sources and microphones were used. The T_{30} and EDT_{15} values, averaged across all source and receiver positions, for both rooms are presented in fig. 3.4.

3.3.1 Salon des Nobles

Acoustic measurements were taken of the Salon des Nobles in June, 2020. A 10 s long sine sweep from 40 Hz to 20 kHz with a sampling of 96 kHz and bit depth of 32 bits used. A Look-



(a) Salon des Nobles. **(b)** Amphitheater.

Figure 3.5 Source (orange dodecahedrons) and receiver (blue circles) positions during the acoustic measurements of the two rooms.

line dodecahedron omnidirectional loudspeaker was used as the sound source while DPA 4006 omnidirectional microphones were used as receivers. Two source positions and six receiver positions were measured (see fig. 3.5a). The positions of the sources and receivers were chosen roughly based on the general positions of musicians and audience in a concert setting. The mean T_{30} across all of the source and receiver combinations at 1 kHz is 1.36 s.

3.3.2 Amphitheater

Acoustic measurements of the amphitheater at Cité de la Musique were taken in April, 2019. A 26 s long sine sweep from 20 Hz to 20 kHz with a sampling rate of 44.1 kHz and bit depth of 16 bits was used. Three Dr. Three (3D-032) dodecahedron omnidirectional loudspeakers were used as sound sources while DPA 4006 omnidirectional microphones were used as receivers. Three source positions, located on stage, along with six receiver positions, located throughout the audience area, were used (see fig. 3.5b). The mean T_{30} across all of the source and receiver combinations at 1 kHz is 1.02 s.

Table 3.1 Clarity values (C_{80}, averaged across the 500 Hz and 1 kHz octave bands) for each source and receiver combination. Source and receiver locations and be seen in fig. 3.5.

	R 1	R 2	R 3	R 4	R 5	R 6
Salon des Nobles						
S 1	0.95	6.73	2.48	1.90	1.66	0.88
S 2	2.24	2.41	3.03	-0.15	0.81	0.93
Amphitheater						
S 1	1.50	1.56	1.81	2.89	1.99	2.40
S 2	1.74	1.71	1.67	2.03	1.64	1.59
S 3	1.42	1.94	2.23	1.49	1.18	1.76

3.3.3 Acoustic comparison

The Salon des Nobles appears to be a fairly typical hall of the baroque era. During this period, "music was usually performed in the music rooms of palaces, most of them rectangular in shape, whose reflecting surfaces were hard and often very richly ornamented" (Beranek, 1992, p. 25). This description is an apt representation of the Salon des Nobles. Additionally, according to studies done in the early 20$^{\text{th}}$-century, the optimum reverberation time for baroque music was found to be 1.6 s in the 500 Hz to 1000 Hz range (Beranek, 1992). The reverberation time in the Salon des Nobles in that frequency range is 1.4 s, suggesting further that it is a fairly typical hall for its time period.

The amphitheater has a moderate reverberation time of around 1 s across all analyzed frequency bands. This reverberation time suggests it may be perceived as "dry" since it is on the lower range of reverberation times of other halls known for their dryness (Beranek, 1962, p. 426). The longer reverberation time of the Salon des Nobles in the 250 Hz to 1000 Hz range may serve as needed acoustical support for baroque instruments which are typically quieter than their modern counterparts (Haynes, 2007, pp. 151–152). Baroque instruments also tend to produce richer overtones compared to modern instruments, translating to increased high frequency energy (Haynes, 2007, pp. 151–152). It is possible that the effect of a shorter reverberation time in the 2 kHz to 4 kHz region in the Salon des Nobles may alter the timbre such that the instrument's sound is perceived as rounder and less harsh. On the other hand, the spectral neutrality of the amphitheater, while preferred for most modern instruments, would not affect the perception of these overtones, resulting in a potentially less pleasing timbre, as

human hearing is particularly sensitive in this frequency range.

However, this flat and moderate reverberation time in the amphitheater may also be an asset. Beranek (1962, p. 45) stated that "[i]n Baroque music the detail is important and no portion of the sound should mask another." The relatively neutral acoustics of the amphitheater may help musicians and audience to hear important details which might be muddied in the Salon des Nobles. Furthermore, the more consistent clarity values, regardless of source or receiver position, in the amphitheater, further support this idea (see table 3.1).

One benefit of many baroque-era rooms was their tendency to sound "intimate" which was due to "the many nearby sound-reflecting surfaces" (Beranek, 1962, p. 45). The Salon des Nobles certainly contains many nearby sound-reflecting surfaces which are likely to impart such an intimate quality to the room's acoustics. This intimate quality may be particularly helpful for weaker-sounding (by modern standards) baroque instruments. Furthermore, this is a particularly desired quality for solo music.

Most of the music performed in the Salon des Nobles was written by composers employed directly by the king and therefore with intimate knowledge of the acoustics of the performance spaces in the Château de Versailles. It has been previously acknowledged that composers have historically adjusted their compositional style to account for the acoustics of the space in which the composition was intended to be performed (Dart, 1954, pp. 56–57). Therefore, one could reasonably expect the acoustics of the Salon des Nobles to work particularly well with instruments and repertoire of the era.

On the other hand, the amphitheater, like many modern halls, was designed as a multi-purpose space, intended to host a variety of ensembles and music genres. The difficulty in designing acoustics for such multi-purpose spaces is apparent, and has been previously documented (Beranek, 1962, p. 481). While some of the acoustic characteristics of the amphitheater may be advantageous, such as the uniformity of sound implied by its stable reverberation time and clarity measures, the acoustics of the Salon des Nobles overall are expected to be more fitting for the performance of solo baroque music.

3.4 Geometrical acoustic models under study

Geometrical acoustic models were created of both rooms using the CATT-Acoustic (v9.2) and TUCT (v2.0e:1.02) software (Dalenbäck, 2016). These models were calibrated to their respective in-situ measurements following the procedure detailed in Postma and Katz (2015). This calibration procedure is an iterative process in which properties of the surfaces within the

model, namely absorption and scattering coefficients, are modified in order to bring certain acoustic parameters, typically T_{30}, EDT_{15}, C_{50}, and C_{80}, to within 1 JND of the measured RIRs.

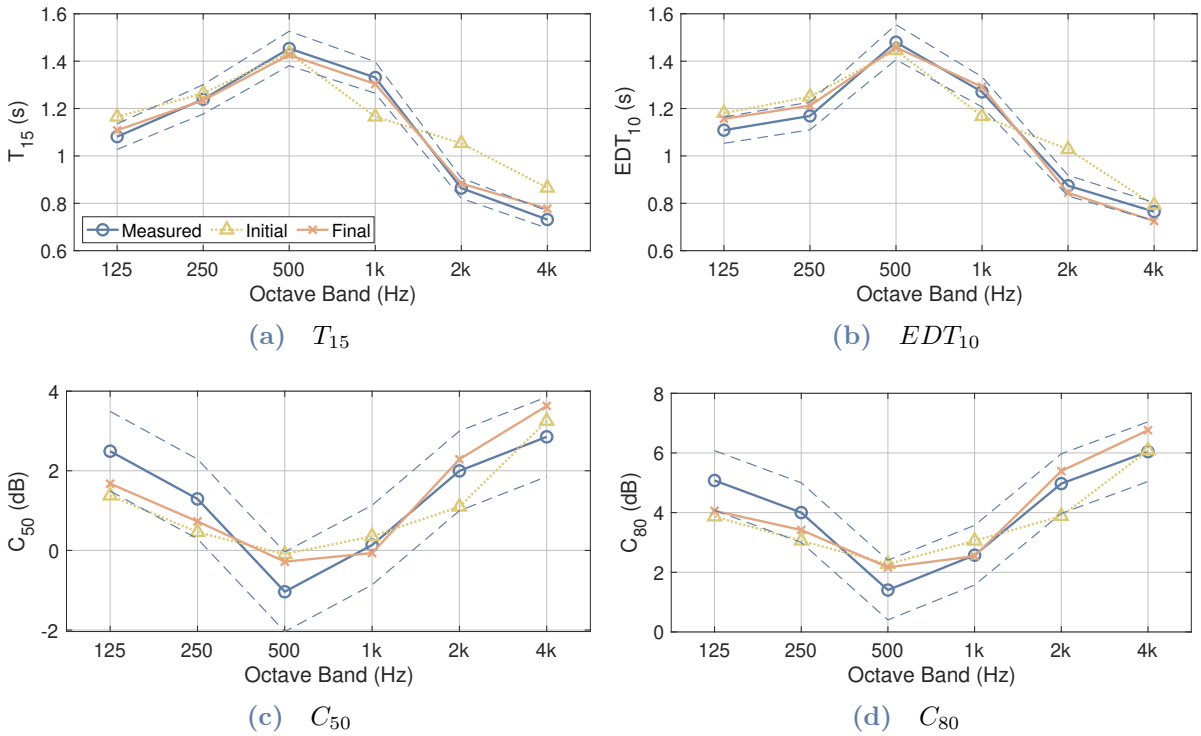
These parameters are used because of their relatively direct relationship to adjustable parameters in the geometrical acoustic models, namely the absorption and scattering coefficients. This also allows for a more accessible calibration procedure since these acoustic parameters are easy to calculate from omnidirectional measurements and do not require special transducers or equipment configuration. While other parameters are more useful in understanding how a musician perceives the room's acoustics, such as ST_{Early} and ST_{Late}, they do not have a predictable relationship to the adjustable parameters in the geometrical acoustic parameters. Therefore, they would not be useful acoustic parameters in the calibration process.

The initial absorption coefficients for the surfaces in the GA model are chosen by manually examining the materials in-situ and choosing best matches in databases of absorption coefficients. Scattering coefficients are applied to a surface as a simplified way to model smaller geometric variations of the surface. These frequency-dependent coefficients are estimated based on the characteristic depth of the surface using the 2D Lambert model in CATT-Acoustic,

$$scatt_{coeff}(f) = 0.5\sqrt{\frac{char_{depth}}{\lambda}} \tag{3.1}$$

where λ is the wavelength and the output is limited to values between 0.10 and 0.99.

The geometry of the model remains unchanged throughout the calibration process. Due to the fact that the GA software implements randomized scattering (Dalenbäck, 2010), there can be some expected variation among repeated runs of the same GA model. To account for this, the run-to-run variation is characterized by comparing the acoustic properties output from 10 runs of the initial model. The properties of the surfaces in the GA model are gradually adjusted, within reason, according to the best estimates of the possible variance of absorption coefficients of the observed materials. The global average of the acoustic properties of five runs are averaged together during each iteration before adjustments to the surface properties are made. This process continues until each acoustic parameter is brought to within 1 just-noticeable difference (JND) of the measured RIRs. At this point, the acoustic parameters of local source-receiver combinations are examined for large differences between the modeled and measured RIRs which may be due to the influence of large nearby surfaces. Adjustments are made as needed while still keeping the global average of acoustic parameters within 1 JND of the measured RIRs.



(a) T_{15} (b) EDT_{10}

(c) C_{50} (d) C_{80}

Figure 3.6 Results from the calibration of the GA model for the Salon des Nobles. Acoustic parameters, averaged across all source and receiver positions, for the initial and final conditions of the calibration process are shown, along with the parameters from the measured RIRs.

3.4.1 Salon des Nobles acoustic model

The geometric model for the Salon des Nobles was created based on physical measurements of the dimensions of the room, taken with a precise laser measure device. The length and width of the room are roughly 9 m each, while the height of the room is about 7 m, yielding an approximate volume of 564 m^3.

During the acoustic measurements of the Salon des Nobles, there was construction work using heavy machinery being done in the courtyard directly outside the room. Consequently, the signal-to-noise ratio (SNR) suffered, particularly in the 125 Hz octave band. This poor SNR made it difficult to reliably calculate T_{30} and T_{20}, in that octave band, so T_{15} was used instead. The other acoustic parameters used in this model calibration were EDT_{10}, C_{50}, and C_{80} as defined by ISO 3382-1 (ISO, 2009) and detailed in section 2.2.1.

The materials and their associated absorption coefficients were mostly drawn from various

databases included in the CATT-Acoustic software. Closest match materials from absorption databases were estimated based on manual inspection of materials in-situ. Most of the walls are covered with a thin tissue of cotton-like material. The ceilings are ornate and appeared to be covered in gilding which has been shown to lower the absorption coefficient of standard materials (Lokki and Pätynen, 2020). This may be one reason for the long reverberation time in the Salon des Nobles relative to its size.

After achieving parameters which fell within 1 JND of the measured values, the individual source-receiver comparisons did not show deviations significant enough to warrant further refinements for this study. Throughout the process, the software used algorithm 1 with approximately 100,000 rays (the suggested number of rays was around 47,000), and first order diffraction was turned on.

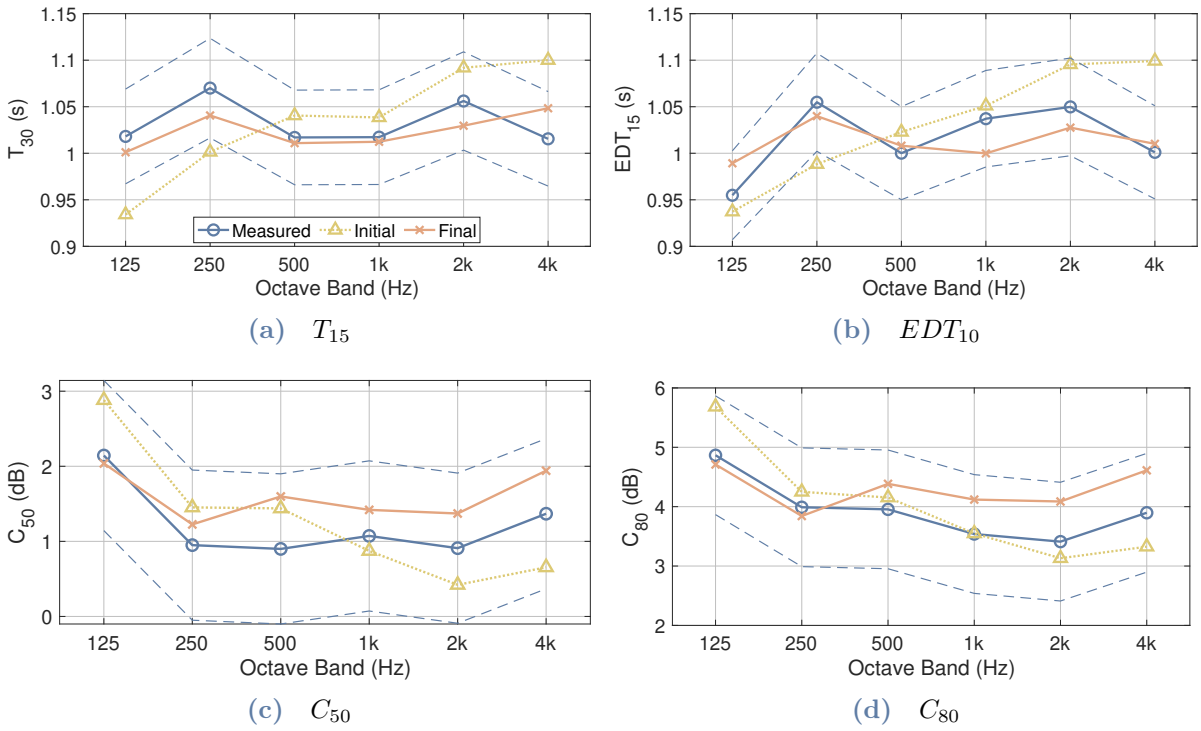
3.4.2 Amphitheater acoustic model

The amphitheater is a roughly fan-shaped, asymmetrical hall with an approximate volume of 1430 m^3. All of its material properties were examined based on manual inspection of the surfaces in-situ.

The SNR of the RIRs recorded in the amphitheater was sufficient to reliably calculate T_{30} and EDT_{15} so these parameters, along with C_{50} and C_{80}, were used in the calibration process. Algorithm 1 was used throughout the process with 50,000 rays (the suggested number of rays was around 20,000), and first order diffraction was turned on.

3.4.3 Predicted acoustic parameters

Several additional acoustic parameters of interest were computed from the calibrated GA models and are reported in fig. 3.8 and table 3.2. JNDs are included in the figures where standardized. The source and receiver were placed on the stage approximately where the musicians performed in the main experiment, separated by a distance of 1 m as recommended in ISO standard 3382-1 (ISO, 2009) for calculating support parameters. This was meant to provide a better approximation of how these parameters might sound to the musicians who took part in the experiment. These parameters were calculated from simulated RIRs as the measurement configuration was intended for calibrating GA models and did not allow for their calculation. It should be noted that the values for G reported in fig. 3.8c are higher than the typical range reported in ISO (2009). The reason for this discrepancy is that, in this particular case, G was calculated from a configuration where both the source and receiver were located on stage. The typical range of values reported in the ISO standards, on the other hand, is derived from a



(a) T_{15} (b) EDT_{10}

(c) C_{50} (d) C_{80}

Figure 3.7 Results from the calibration of the GA model for the amphitheater. Acoustic parameters, averaged across all source and receiver positions, for the initial and final conditions of the calibration process are shown, along with the parameters from the measured RIRs.

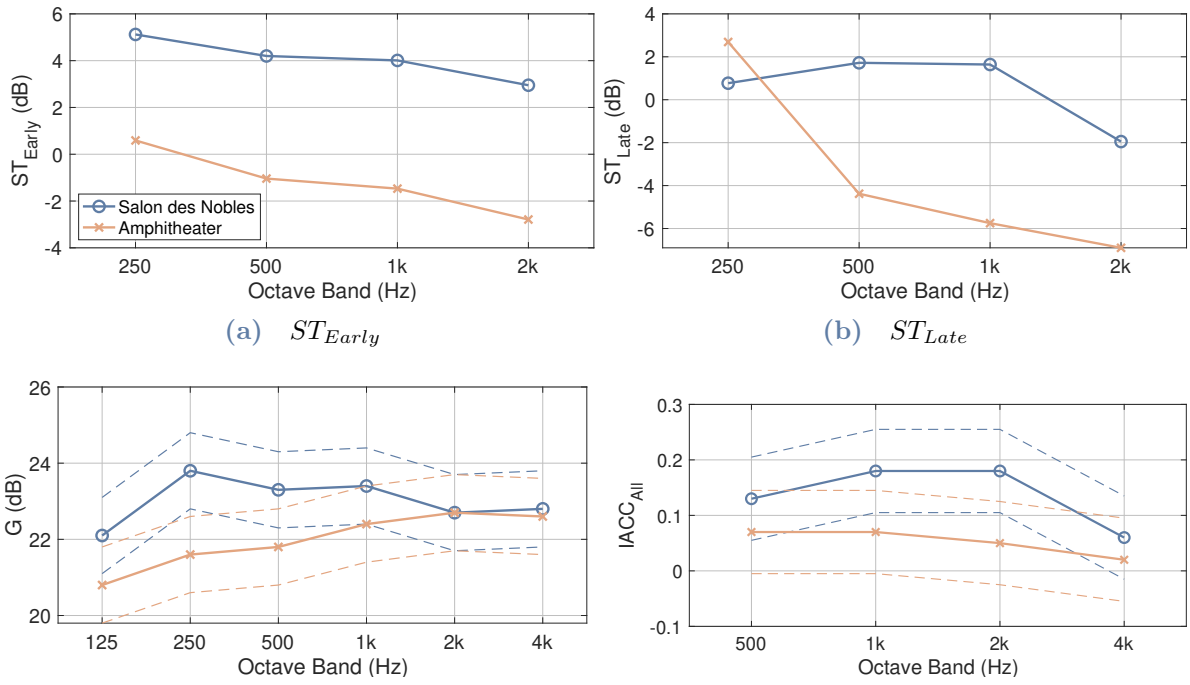
more common configuration where the source is on stage and the receiver is situated in the audience area.

3.5 System calibration

The following sections describe the calibration of the auralization system in terms of overall level (section 3.5.1) and spectral balance (section 3.5.2).

3.5.1 Level calibration

Conveying simulated acoustics at the proper level must be done carefully so as not to artificially amplify or attenuate the acoustics. Several procedures have been developed for this purpose (Amengual Gari et al., 2019; Schärer Kalkandjiev, 2015; Laird, Chapman, et al., 2012). This



(a) ST_{Early} (b) ST_{Late}

(c) Sound Strength (G) (d) Inter-aural cross correlation ($IACC$)

Figure 3.8 Additional acoustic parameters calculated from the calibrated GA models of the two rooms. JNDs, where known, are shown as dashed lines.

calibration step should be used to achieve a relatively reliable starting level rather than to determine an absolute fixed level. Minor adjustments may be expected based on various aspects of the configuration, such as the instrument's directivity pattern and the exact placement of the microphone.

For the EVAA system, a slightly modified version of the procedure outlined in Laird, Chapman, et al., 2012, was used. This method compares the energy in an impulse response to the direct sound as defined within certain time intervals. An acoustic parameter, first defined by Gade (1989a) provides a more formal definition, termed stage support. It exists as both early support (ST_{early}, see eq. (2.10)) and late support (ST_{late}, see eq. (2.11)). Gade described this as the parameter primarily responsible for allowing musicians to hear themselves on stage (ST_{early}) or to hear the hall from the stage position (ST_{late}). The calibration procedure relies on comparing the ST_{late} between the VAE and the real acoustic space to produced a gain coefficient by which the simulated acoustics can be amplified or attenuated. As noted earlier, the configuration required for calculating ST_{late} (a source and receiver separated by 1 m) was not

Table 3.2 Additional acoustic parameters computed from the calibrated GA models of the two rooms.

	Salon des Nobles	Amphitheater	JND
LF_{early}	0.17	0.10	0.05
Bass Ratio	0.83	1.03	n/a
Treble Ratio	0.57	1.02	n/a

included during the measurements of the halls used in this study. To rectify this, the ST_{late} parameter was derived from the calibrated GA models described in section 3.4.

The procedure is as follows. A set of beam impulse responses (IRs) is produced from the GA model corresponding to the proper source and receiver positions needed to calculate ST_{late}. The ST_{late} parameter is then measured through the EVAA system. A loudspeaker (Genelec 8030a) and omnidirectional microphone (DPA 4060) a few centimeters away are used to send an exponential sine sweep through the system. The plug-in must have the set of RIRs described above, along with a specific instrument directivity pattern. The output is then captured by an omnidirectional microphone (DPA 4006) positioned 1 m away, in the same configuration as the GA model. This sweep is then deconvolved to provide the IR, which can be used to calculate the ST_{late} parameter. This measure is then compared to the same parameter calculated from the RIR produced by the GA model and a difference, in dB, is calculated between the two. This gain factor is then applied to the output of the system during experiments. It is necessary to perform this level calibration procedure for each different instrument directivity, room, or location within a room.

3.5.2 Loudspeaker equalization

In addition to calibrating the overall level of the auralization system, an equalization was applied to each of the 32 loudspeakers in order to ensure an optimum frequency response at the sweet spot. The loudspeakers were equalized using 12^{th}-order biquad filters. A sine sweep was recorded individually from each speaker using a DPA 4006 omnidirectional microphone situated in the center of the loudspeaker array. IRs were created from these sine sweeps and used to create inverse compensation filters using a MATLAB script. These filter coefficients where then implemented in the main Max patch.

3.6 Visualization

As a complementary component to the auralization, an immersive, adaptively rendered visualization of each hall was also included. While there is some disagreement as to whether including visual stimuli may help or hurt a study investigating musicians' performance (see section 3.6.1), with the goal towards creating a totally immersive interactive environment, this visual component was included in the main experiment.

3.6.1 Justification

The decision to include complementary visual stimuli is not inconsequential, and there is no consensus as to whether a visual component is necessary or even helpful in studies investigating the role of acoustics on music performance. If the research question is simply to better understand the effect of acoustics alone on music performance then it may be justified to exclude visuals in order to isolate the effects of acoustic changes from other influencing factors. This was a justification for the VAE used in Luizard, Brauer, et al. (2019), where the authors wrote that they were able to avoid "the confusion with other influencing factors such as the visual impression, audience reactions, or the architectural design of the space." Additionally, some previous research has suggested that visual feedback probably plays a relatively unimportant role on musicians' performance since it is fairly common for musicians to close their eyes while performing (Repp, 1999b). Furthermore, Ciaramitaro et al. (2017) describes how decreasing one's visual load can free up neural processing to be devoted to audio stimuli, suggesting that a study excluding visual stimuli may yield stronger results by making it easier for the musician to focus on acoustic changes.

However, if the research question seeks to understand how musicians will realistically react to changes in acoustics then depriving them of visual stimuli may be detrimental since it is impossible to separate acoustics from the physical space in real performance scenarios. Indeed, Schärer Kalkandjiev and Weinzierl (2015) suggested that the effect of the visual impression of the hall may strengthen the effects of the acoustic changes, stating that "the auditory and visual impressions taken together might have a stronger impact than the isolated auditory information alone." In reviewing that study, Kob et al. (2020) stated, "it seems the absence of visual information about the concert halls in the laboratory study did not aid the musicians' concentration on adjusting to the room acoustics. Instead, much attention was drawn by the effort to get a mental image of the simulated rooms, as interviews revealed. Furthermore, visual and acoustical properties of rooms usually covary so that they may have a stronger effect as an

(a) (b)

Figure 3.9 A side-by-side comparison of the (a) Salon des Nobles and (b) its virtual rendering.

entity." Given that performance changes are relatively subtle, such a strengthening effect may be highly beneficial. In another study examining the effect of acoustics on musicians which used an auralization but included no complementary visual stimuli (Canfield-Dafilou et al., 2019), the researchers found that some of the musicians were "distracted by the incongruence between the sound of the auralization and the visuals of the recording studio."

Lastly, there is some evidence that visual stimuli can have an impact on auditory perception. Stein and Meredith (1993) stated that "perceptual information from one sense, such as vision, influences evaluation and perception of information in other senses, such as hearing." However, the precise nature and magnitude of this interaction has not been well-studied. For example, while Postma and Katz (2017) found that the perception of loudness and sound source distance were influenced by visuals, Schutte et al. (2019) and Salmon et al. (2020) found that visual stimuli had no statistically significant effect on the perception of sound spaces.

Taking these points into consideration, it was decided to include a visual component to make a multimodal immersive environment. This means that the while the primary research question remains how the acoustics affects music performance, other factors, such as the visual impact, can be taken into consideration. For one, there are not many studies that have done this so any findings, positive or negative, could be valuable. Second, this study would be an important step towards total immersion and realism in virtual reality (VR) environments and could provide valuable insights into the development of such systems in the future.

3.6.2 Creation of visual models

The visual models were created in Blender and were based on photogrammetry captured in-situ. The model of the Salon des Nobles (see fig. 3.2) was created by a team of students of the *licence professionelle*, "Patrimoine, Visualisation et Modélisation 3D" at CY Cergy Paris University. The model for the amphitheater (see fig. 3.3) was created by a contracted professional. The assets for both of these visual models were delivered as Blender files. These were then exported as *.fbx* files and imported into the Unity 3D engine where the materials, lighting, and shaders were all updated for use in the realtime rendering engine. It was originally intended to use the high-definition render pipeline (HDRP) in Unity, but after encountering several issues with improper rendering of certain materials, it was decided to revert back to the original renderer. Unity was used rather than Blender, since its real-time renderer is more flexible than that of Blender, allowing for such possibilities as adaptive rendering using an existing framework (see section 3.6.4).

3.6.3 Physical framework

There are several visual rendering systems with the two most common being head-mounted displays (HMDs) and cave automated virtual environment (CAVE) systems. Shiratuddin and Sulbaran (2006) compared three different visual rendering systems: an HMD, a CAVE, and an immersive WorkBench. The researchers found their suitability to present 3D visuals was roughly equivalent. Another comparison of visual rendering systems (Kim et al., 2012) found that an HMD display was more likely to evoke a negative emotional response than a CAVE system. Another study looking into the impact of visual rendering systems on assessments of auralizations (Thery et al., 2017) found that "the choice of VR visual rendering system had little impact on the subjective evaluation of selected acoustical attributes."

Given the above findings, the choice on which visual rendering system to use should be based on other practical considerations. While an HMD may provide better immersion, it would likely interfere with a musician's ability to perform to an unacceptable degree. Given that changes in playing style are relatively subtle and can be influenced by a number of factors, the virtual system should be as unobtrusive as possible. Therefore, a CAVE approach was chosen to render the visuals. This system consisted of four screens: one front screen (3.32 m x 2.22 m), two side screens (1.04 m x 2.22 m), and a floor screen (3.32 m x 1 m).

The screen was made of an acoustically transparent material so as to minimize its effects on the signals from loudspeakers behind it. The frame was made of modular aluminum profiles

(a) (b)

Figure 3.10 A side-by-side comparison of the (a) amphitheater and its (b) virtual rendering.

which are about 4 cm x 4 cm wide.

3.6.4 Adaptive rendering

Real-time adaptive visual rendering, a feature which allows the image to vary in a natural way according to the position and movement of the participant was implemented using the same motion capture system which was used to implement dynamic directivity.

This feature was achieved using the UniCAVE system (Tredinnick et al., 2017). This is a Unity plug-in which links the virtual camera in Unity to the position of the participant using the motion capture system. A virtual screen setup mimicking the physical setup in the room is added to Unity which allows for the proper transformations and perspective shifts based on the movements and position of the viewer in real time. The virtual screen must be carefully aligned with the physical screen first by projecting anchor points at the edges and vertices of the screens, in order for the transformations to be properly applied.

The motion capture device is fixed to the participant's head using a headband. For vocalists and certain instruments (such as the flute) this is the same device used to monitor their orientation for implementing dynamic directivity. In other cases, two devices are needed, one to monitor the participant's position for adaptive visual rendering, and one to monitor the orientation of their instrument, for dynamic directivity.

3.7 Comparison with other virtual acoustic environments

Several other VAEs have been developed specifically for studying the effect of acoustics on solo musicians (Ueno, Yasuda, et al., 2001; Brereton, 2014; Schärer Kalkandjiev, 2015; Amengual Gari, 2017). Most of these VAEs use measured SRIRs as the basis for their auralization whereas the EVAA system uses RIRs derived from a calibrated GA model. One other study used a VAE derived from a GA model (Schärer Kalkandjiev, 2015) but no effort was made to align the acoustic parameters from the model with acoustic parameters of measured RIRs. In a prototype of the VAE used in Brereton (2014), an auralization based on GA model which had been optimized against in-situ measurements (Foteinou et al., 2010) was compared to an auralization using measured RIRs, and most of the participants (seven out of eight) preferred the VAE using measured RIRs. This led to a decision to use measured SRIRs for the main implementation of the VAE in that study.

One major benefit to using RIRs from a calibrated GA model is that it allows for flexible source and receiver configurations whereas VAEs relying on measured RIRs are constrained to the arrangements which were recorded. It also allows for the possibility of studying different room configurations by modifying the geometry or the surface properties of the GA model. Because of the calibrated nature of the GA model, one can be confident that the resulting acoustics from the modified model are reasonably close to reality.

The EVAA system is the only VAE of those mentioned to include any kind of dynamic directivity. Additionally, most of the other VAEs used only a generalized version of instrument directivity, typically relying on similarities of the directivities of the instrument under study and the source used when taking measurements (Brereton, 2014; Amengual Gari, 2017). One disadvantage to systems that rely on such setups is that they are inflexible when it comes to using instruments with directivity characteristics that do not resemble loudspeakers.

Most of the mentioned VAEs use a loudspeaker array as the playback medium for the virtual acoustics since most musicians find headphones bothersome when performing. However, Schärer Kalkandjiev (2015) used a binaural-over-headphones approach with head-tracking. The EVAA system is flexible and can switch between these two playback approaches fairly quickly and easily.

The spatial resolution of the EVAA system is also superior to many other systems, using a 32-channel loudspeaker system fed with a 3$^{\text{rd}}$-order Ambisonic RIR. Most previous studies used 1$^{\text{st}}$-order Ambisonics or some similar spatial recording method as the basis for their system. However, the perceptual significance of this higher resolution in studies from a performer's

perspective, if any, is not yet known.

3.8 Preliminary experiment

In order to evaluate the auralization system from the point of view of musicians, a preliminary study was designed with four professional singers performing in a VAE of the Cathedral of Notre-Dame de Paris. This project took place as part of the The Past Has Ears at Notre-Dame (PHEND) project[2]. One primary difference with the setup of this auralization architecture is that, due to the long reverberation time of Notre-Dame and the duration required to calculate such a long set of RIRs (including 20 beams for implementing dynamic directivity) for auralization, only the first second of the RIR was calculated. The remainder of the RIR was approximated using a feedback delay network (FDN) tuned to the first part of the RIR using the IEM FdnReverb[3] that takes higher order ambisonics (HOA) as input. Another key difference is that this study did not include any visual component.

3.8.1 Study participants

The four singers who participated in the experiment were members of a professional medieval choir ensemble which focuses primarily on historically informed performance of music from the 12^{th} and 13^{th} centuries. The musicians were chosen partly due to their familiarity with the acoustics of Notre-Dame, in which they all had sung, since part of the aim of this preliminary study was to perceptually validate the virtual replication of Notre-Dame's acoustics.

3.8.2 Experimental overview

The musicians individually sang excerpts of medieval repertoire with which they were previously familiar. They spent 10 to 15 minutes interacting with the VAE. At the beginning of the session, the subjects were asked about the overall sound level of the virtual room acoustics in order to assess the effectiveness of the level calibration procedure (see section 3.5.1). The gain was then adjusted to their preferred level before continuing. They were given the opportunity to sing and move freely and also interact with the VAE by voicing plosive, fricative, and sibilant sounds outside of a musical context, in order to better explore the dynamic directivity component of the auralization component. The participants experienced the VAE over loudspeakers and also through open-back headphones (Sennheiser HD 650), in order to compare these two

[2] http://phend.pasthasears.eu
[3] https://plugins.iem.at/docs/plugindescriptions/

reproduction methods. They provided feedback throughout the process, and also completed questionnaires related to the perceptual validity of the experience.

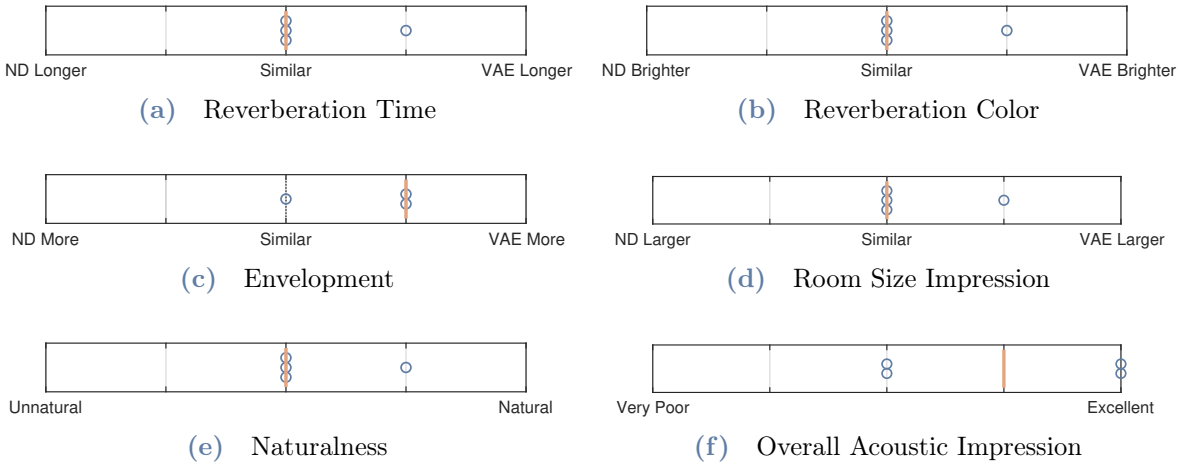
While the participants were not singing, they participated as listeners in another room, in order to provide insight into the effectiveness of the auralization from the point of view of a nearby listener position. This listener position was chosen to approximate a position that would be typical in an ensemble setup, about 1 m from the singer. The auralization of this position was created in same way as the singer's but with the direct sound included. The listeners were situated in a separate room so they could not hear the singer directly. The listeners also filled out questionnaires related to their perceptual experience.

The subjects were asked to rate their aural experience singing in the VAE in comparison to their previous experience singing in Notre-Dame within the categories of *reverberation time*, *reverberation color*, *envelopment*, and *impression of room size*. For example, if they rated the "envelopment" as more enveloping, they would be stating the VAE felt more enveloping than Notre-Dame. The questions were administered on a 5-point scale with the middle point indicating a "similar" rating with the remaining points labeled appropriately in their respective categories. In addition to the above categories, the participants were asked to rate the virtual acoustics in terms of *naturalness* and *overall acoustic impression* on a 1 to 5 scale. The same questionnaire was given to the subjects when they participated as listeners. The questionnaires were developed in english then translated to french by a native speaker to be administered to the participants. The results are reported here in english.

3.8.3 Results

It should be noted that this study included only four participants who were basing their ratings on a comparison with experiences which occurred at least two years prior, so any conclusions drawn from the results should take this into consideration. The results of the singers (fig. 3.11) and the listeners (fig. 3.12) indicate that the VAE was evaluated as being fairly similar to Notre-Dame.

In terms of naturalness, on a scale from 1 (unnatural) to 5 (natural), the median rating given by the singers and listeners was 3. The median overall acoustic impression, on a scale of 1 (very poor) to 5 (excellent) given by the singers was a 4 while for the listeners it was 3. This difference in rating between the participants as singers compared to the participants as listeners may be explained by the cognitive load required to concentrate on singing which likely makes an assessment of the acoustics more difficult. Additionally, the context for participants as listeners—seated at a table—was not the most ecologically valid condition for ensemble



(a) Reverberation Time (b) Reverberation Color

(c) Envelopment (d) Room Size Impression

(e) Naturalness (f) Overall Acoustic Impression

Figure 3.11 Individual responses (circles) and medians (lines) to the singer questionnaire from the preliminary experiment.

listening.

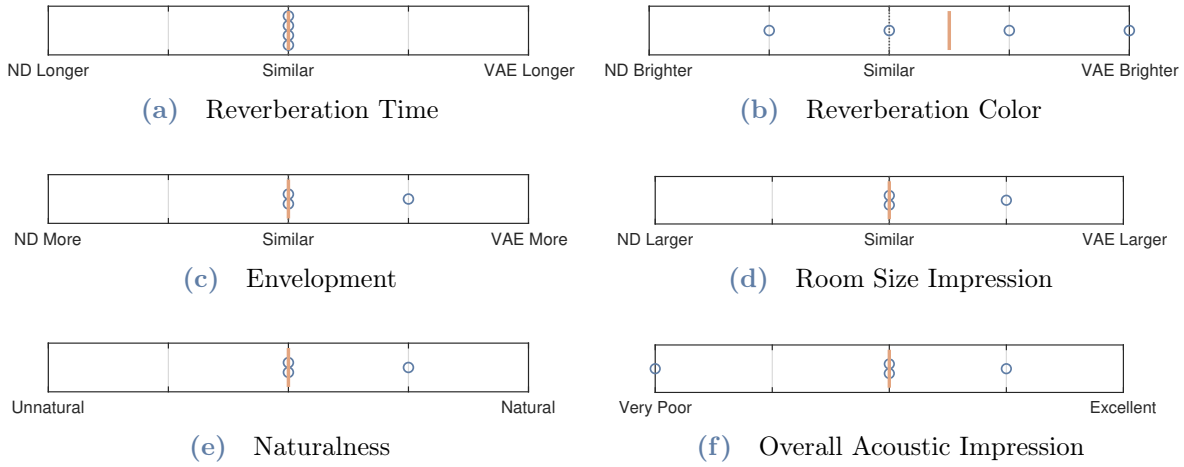
When asked explicitly about the dynamic directivity component, three out of four singers said they were unable to notice any difference in the sound based on their orientation. One singer noted that they noticed a slight difference when actively rotating and that it added to their sense of envelopment.

Two singers preferred the virtual acoustics through loudspeakers while one singer preferred it through headphones. The fourth singer had mixed feelings, saying that the headphones, while making the acoustics sound more realistic, also made it difficult to hear himself.

Only one singer asked for the level to be changed (an increase of approximately 2 dB) when using loudspeakers as reproduction method. When using headphones, three singers requested a similar change. This indicates that the level calibration procedure is fairly effective, especially for loudspeaker reproduction, though some improvement is possible.

One notable comment made by the participants had to do with the coloration of the system. All participants noted an unnatural and metallic timbre present in the system, indicating an excess of high frequency energy in the auralization. These comments were despite a fairly broad consensus among participants in the questionnaires that the reverberation color was similar between the VAE and Notre-Dame. However, it has been previously found that singers in VAEs have difficulty assessing timbre (Brereton, 2014).

In examining the system due to this feedback, it was found that there was a flaw in how the FDN was being calibrated to the initial portion of the reverb (from the GA model). This error



(a) Reverberation Time (b) Reverberation Color

(c) Envelopment (d) Room Size Impression

(e) Naturalness (f) Overall Acoustic Impression

Figure 3.12 Individual responses (circles) and medians (lines) to the listener questionnaire from the preliminary experiment.

caused reverberation time of high frequencies to be longer than they should have been which could have been a cause of this high-frequency coloration. However, because the auralizations in the main experiment (see chapter 4) did not require an FDN, no follow-up study was carried out in the course of this thesis to confirm that this was the primary cause of the coloration.

3.8.4 Discussion

This preliminary study showed that the VAE was able to convey the acoustics of Notre-Dame with fairly high accuracy according to singers with experience performing in the cathedral. However, the applicability of these findings to the main experiment (see chapter 4) in this thesis may be limited for several reasons. First, the auralization architecture is slightly different since in the main experiment, no FDN was used. Second, the acoustics under consideration are very different (a large cathedral in the preliminary study and small chamber halls in the main experiment). Lastly, the type of music under study in the main experiment is very different from the music examined in the preliminary study. For these reasons, it is difficult to know to what extent the findings from this preliminary study can be applied to the main experiment.

The dynamic directivity component was not very evident, even in the extreme cases of a participant spinning while vocalizing transient sounds. This may have been partly due to the long reverberation time, which could mask the aural cues necessary for hearing such a difference. Additionally, most participants did not deviate from their normal singing posture, other than slight head movements raising questions as to whether a dynamic directivity component is even

necessary in scenarios such as this.

The participants had a slight preference for loudspeaker playback over headphones. Despite the usage of open-back headphones, some participants complained that the selected headphones muted their own voice. This aligned with previous studies that found that musicians disliked wearing headphones while performing (Brereton, 2014; Schärer Kalkandjiev, 2015).

One criticism, which was shared by all participants, was the problem of high-frequency coloration. While an issue with the FDN was identified which was thought at the time to be the cause of the coloration, similar complaints were later found in response to the main experiment in which no FDN was used (see section 4.3).

3.9 Summary

This chapter detailed the design and development of the EVAA system to aid in studying the effect of room acoustics on musicians. The EVAA system in general was described, as was its specific implementation for the experiments performed within the framework of this thesis. The room selection process for this study was described as well as the acoustic measurements and parameters of the two rooms selected. The acoustics of the two rooms were compared in the context of other halls and the expected impact on the performance of baroque music. GA models of these two rooms were created and calibrated, and these models formed the basis of the auralization system. Additional novel elements of the system were described including the dynamic directivity component and the immersive visualization in addition to justifications for including these features. To put the EVAA system into perspective, it was compared to other auralization systems which have been used to study the effect of room acoustics on musicians. Finally, a preliminary experiment was described which was designed to test the system for use by musicians. The experiment was relatively successful although it did reveal some potential issues with the system such as a high-frequency coloration, and the difficulty in assessing dynamic directivity. It was discussed that these issues may have been the result of elements which were particular to this acoustic simulation which utilized an FDN.

This information is essential to the next chapter (chapter 4) which describes the main experiment of the study in which musicians performed in two real acoustic settings and their virtual counterparts as implemented by the EVAA system.

Chapter 4

Experiment

In order to shed light on the impact of acoustics on the performance of historical baroque music, an experiment was designed in which musicians performed the same repertoire in various acoustic settings including both historical and modern spaces. As previously discussed, there are advantages and disadvantages to in-situ studies and those which use virtual acoustic environments (VAEs). This study used both approaches—enlisting musicians to perform in real rooms and in virtual environments simulating these same rooms. One of the primary goals was to better understand if the acoustics of the baroque-era room facilitated the historically informed performance of that kind of music. In this chapter, this questions is primarily investigated through a set of questionnaires given to the participants after each session. In addition to the questionnaires, different strategies of objective music performance analysis were pursued which are detailed in chapters 5 and 6. Lastly, with the goal of understanding the perceptual salience of the musicians' performance changes, a listening test was performed which is detailed in chapter 7.

4.1 Experimental design

In this study, 10 musicians, all specializing in historically informed baroque performance practice, played in four different conditions: two real halls and two virtual simulations of these

Table 4.1 Schedule of sessions with the date and order that all participants experienced the different settings.

	29 March Amphitheater	2 April Virtual	6 April Salon des Nobles	14 April Amphitheater	15 April Virtual	16 April Virtual
Viol 1		✓	✓	✓		
Viol 2	✓	✓	✓			
Viol 3		✓	✓	✓		
Viol 4	✓		✓		✓	
Flute 1			✓	✓		✓
Flute 2	✓		✓			✓
Flute 3		✓	✓	✓		
Theorbo 1	✓	✓	✓			
Theorbo 2	✓		✓			✓
Theorbo 3			✓	✓	✓	

halls. The halls were chosen to provide an acoustic space which was historically appropriate to baroque music and a contrastive modern hall. One hall is part of the Château de Versailles and hosted concerts of the baroque era and the other hall was built in the late 20^{th} century. These halls and their acoustics are discussed in additional detail in section 3.2. The design and calibration of the immersive virtual environments based on these rooms is described in detail in chapter 3.

The musicians performed several pieces in each setting, repeating each set of pieces two to four times, as time allowed. These repetitions were interleaved with other participants and were not consecutive. Each musician participated in three separate sessions, one session each was devoted to the real halls and both virtual halls were experienced on the same day with a break in between. The time between each session was ten days or less and the order for each musician was randomized as best as possible (see table 4.1 for details on the order and timing of each session for all participants). The protocol was approved by the Research Ethics Committee (*Comité d'Éthique de la Recherche*) at Sorbonne University (approval #: CER-2021-039).

Performances were recorded with a cardioid microphone (AKG C414) positioned 1 m away, directed towards the instrument. The microphone was positioned slightly to the left of the musician to avoid reflections from the music stand. This position was chosen to maximize the direct-to-reverberant ratio of the recorded performances while avoiding physical interference with the musician as they performed.

After each session, participants responded to a number of questionnaires regarding their impression of the acoustics and its potential influence on their performance, as well as their assessment of the virtual environment (see section 4.1.3).

4.1.1 Musicians

There are many factors which were considered when enlisting the musicians for this experiment. Békésy (1968), Ternström (1989), and Bolzinger and Risset (1992) found that experience played an important role in how musicians adapted to changes in acoustics. In general, less experienced musicians will be more occupied with simply meeting the technical demands of the composition rather than on optimizing the performance for the space. Therefore, it is important to use musicians with sufficient professional performance experience. Furthermore, it is also essential to record the level of professional experience of each musician and to take this into consideration during the analysis stage.

Another element under consideration was whether to study soloists or small ensembles. During the baroque era, solo performances on non-keyboard instruments were somewhat rare, as musicians were usually accompanied by *basso continuo* parts. If the primary concern of this study was historical plausibility or authenticity then a small ensemble would have been desirable. This may have also made it easier for musicians to hear the room's acoustics since their own direct sound would not play as dominant a role in masking the acoustics in an ensemble setting. However, considering this study is already examining several novel components, adding another layer of complexity of testing a small ensemble which would involve separating the influence of the musicians on each other from external factors, would have been inadvisable. Additionally, there are relatively few ensemble studies in the acoustics-performance interaction literature, meaning the results would be more difficult to contextualize. Of course, this comes at the cost of the repertoire not being the most representative of the genre, but it was deemed justifiable for the sake of simplifying the already complex problem.

Another factor to consider is that different instruments may require different strategies to respond to changes in room acoustics. For example, string players, which must contend with the resonant bodies of their own instruments, may react differently to changes in acoustics than a wind player. Furthermore, the expressive mechanisms of each instrument are different. For example, expressive gestures requiring slight pitch deviations, such as vibrato, are possible on some string instruments but not on wind instruments. For these reasons, it is important to include a variety of instruments in order to better explore the variety of possible responses.

This study was done with 10 professional musicians trained in historically informed per-

	Age	Yrs Exp.	Handedness	Vision loss	Hearing loss	Sex
Violist 1	61	40	Right	No	No	Female
Violist 2	53	31	Right	No	No	Male
Violist 3	61	41	Right	No	No	Male
Violist 4	22	3	Right	No	No	Female
Flutist 1	29	7	Right	No	No	Female
Flutist 2	62	40	Right	No	No	Male
Flutist 3	28	1	Right	No	No	Female
Theorbist 1	49	25	Right	No	No	Male
Theorbist 2	32	6	Right	No	No	Male
Theorbist 3	31	13	Left	No	No	Female

Table 4.2 Participant information.

formance (HIP). Three different instruments were represented: four violists da gamba, three transverse flutists, and three theorbists. These instruments were chosen because of the availability of solo repertoire, and their suitability to french baroque musical contexts. The average age of the participants was 42.8 years (SD: 15.9). The average number of years of professional experience reported was 20.7 (SD: 16.5). Half of the participants were female and half were male. There was no self-reported hearing loss among the participants, and one participant was left-handed. They were compensated for participating in the study. Table 4.2 contains basic information reported by the participants.

4.1.2 Repertoire

As discussed in section 2.3, the strategies used by musicians to adapt to changes in acoustics may partially depend on the musical content. It is therefore important that the experimental design includes compositions which showcase the variety of playing styles within the baroque era.

Each musician within each instrument class performed the same repertoire. This approach makes it possible to assess the inter-musician variability which is essential for contextualizing the effect of the room acoustics on their playing.

The primary disadvantage is that the quality of the performance may suffer, since the musicians may not be sufficiently familiar with the compositions. As previously discussed, if the musician is preoccupied with the technical demands of the piece, their reaction and adjustment to acoustic changes will not be as evident. This is one of the main reasons that most previous

studies have left the decision of repertoire up to the musicians themselves, advising them to choose repertoire with which they are intimately familiar. Only one study reviewed (Ueno, Kato, et al., 2010) had all the musicians perform the same repertoire. In anticipation of this potential disadvantage, the compositions were chosen partially based on their ability to be learned quickly and the musicians were given the repertoire in advance in order to prepare.

The repertoire was chosen with the assistance and advice of musicologists. The pieces chosen for each instrument are discussed below and can also be found in table 4.3 along with a few attributes describing them. These composers all had an important relationship with musical life at the Château de Versailles as documented in Baumont (2007).

4.1.2.1 Viol da gamba

Marin Marais was one of the most prolific composers of French viol music and a musician to the royal court at Versailles in the late 17th century and early 18th century. One of his better known pieces of music, *Les folies d'Espagne* was published in 1701 and is made up of 32 couplets. The 12th, 13th, and 18th couplets were chosen for this study as they showcased a variety of styles. Although these couplets were often performed with a continuo, they were performed solo for this study.

4.1.2.2 Flute

Jacques Hotteterre was a composer, performer, and maker of wind instruments. He served as a musician to the royal court of both Louis XIV and Louis XV. Three preludes were chosen from his *L'art de Préluder* from 1719 which showcase different tempos and styles: 4ème Prelude G. Re, Sol, 3ce. Majeur (Animé), 4ème Prelude in D. La, Re, 3ce Majeure (Gravement), and 1ère Prelude in C. Sol, Ut, 3ce Mineure (Lentement). The last piece, Lentement, contains a coda which only some musicians took, so each recording was edited to remove the codas in order to unify the performances. These pieces were likely often performed with a continuo, however, for this study, the pieces were performed solo.

4.1.2.3 Theorbo

The musical selections for the theorbo were written by Robert de Visée. De Visée was active as a composer and performer of the guitar, lute, and theorbo in the late 17th and early 18th century and an important musical presence at the Château de Versailles during the time of Louis XIV. A Prelude and a Gigue, composed by de Visée, was chosen for this study. The

Table 4.3 List of musical pieces chosen for the experiment along with a few basic attributes. *Tempo* is an approximation of the average tempo across all performances.

Instrument	Title	Time Sig.	Tempo	# Bars
Viol	Couplet 12	3/4	♩ = 70	16
	Couplet 13	6/8	♩. = 60	16
	Couplet 18	3/4	♩ = 100	16
Flute	Animé	2/2	𝅗𝅥 = 80	14
	Gravement	4/4	♩ = 60	13
	Lentement	4/4	♩ = 100	9
Theorbo	Gigue A	3/4	♩ = 120	8
	Gigue B	3/4	♩ = 120	16
	Prelude	4/4	♩ = 60	20

Gigue has an A and a B section, both of which have optional repeats, and while the musicians were advised not to take the repeats, they were not consistent in their performance. Some of the music analysis performed (see chapter 5) requires each piece to have the same form, so each performance was edited to include only one section. For this reason, all analysis will treat the A and B section of the Gigue as separate pieces.

4.1.3 Questionnaires

The acoustic rating questionnaire administered to each musician after each session was adapted from the Stage Acoustic Quality Inventory (STAQI), designed for subjective evaluations of stage acoustics by musicians (Schärer Kalkandjiev and Weinzierl, 2018). Each of the terms were presented in a semantic differential format on a 5-point line with their extremes (e.g. dry—reverberant) labeled. Additionally, there were open-ended questions about the performers' experience, including any conscious adjustments they believed they made to their performance. These questionnaires (and their french translations) can be found in tables A.1 to A.3.

Another questionnaire, which will be referred to as the "virtual questionnaire" was given to each participant after each virtual environment. This questionnaire asked the participants to rate the virtual environment according to a number of statements such as, "what I saw in the virtual environment was realistic/natural" on a 5-point scale from "strongly disagree" to "strongly agree" (see tables A.4 and A.5). The goal of this questionnaire was to evaluate the overall impression of the virtual environment and identify areas that most needed improvement. Additionally it could shed light on any significant differences between the two virtual

environments.

At the end of the experience, each participant was given a final questionnaire, the "comparison questionnaire," which asked them to compare the similarity of their experiences between the two real rooms and also to compare the similarity of their experiences between the real and virtual environments. This was administered on a 5-point scale from "very different" to "similar" (see table A.6).

Lastly, a "presence questionnaire" was administered to the participants at the end of the virtual session (this questionnaire was not made available due to copyright reasons). Presence is often used as a quality measure to evaluate virtual experiences. It is defined as the subjective experience of "being there" (Barfield et al., 1995). This questionnaire was based on the ITC-Sense of Presence Inventory (ITC-SOPI) (Lessiter et al., 2001) and was used to gauge the validity of the virtual system across four factors: sense of physical space, engagement, ecological validity, and negative effects.

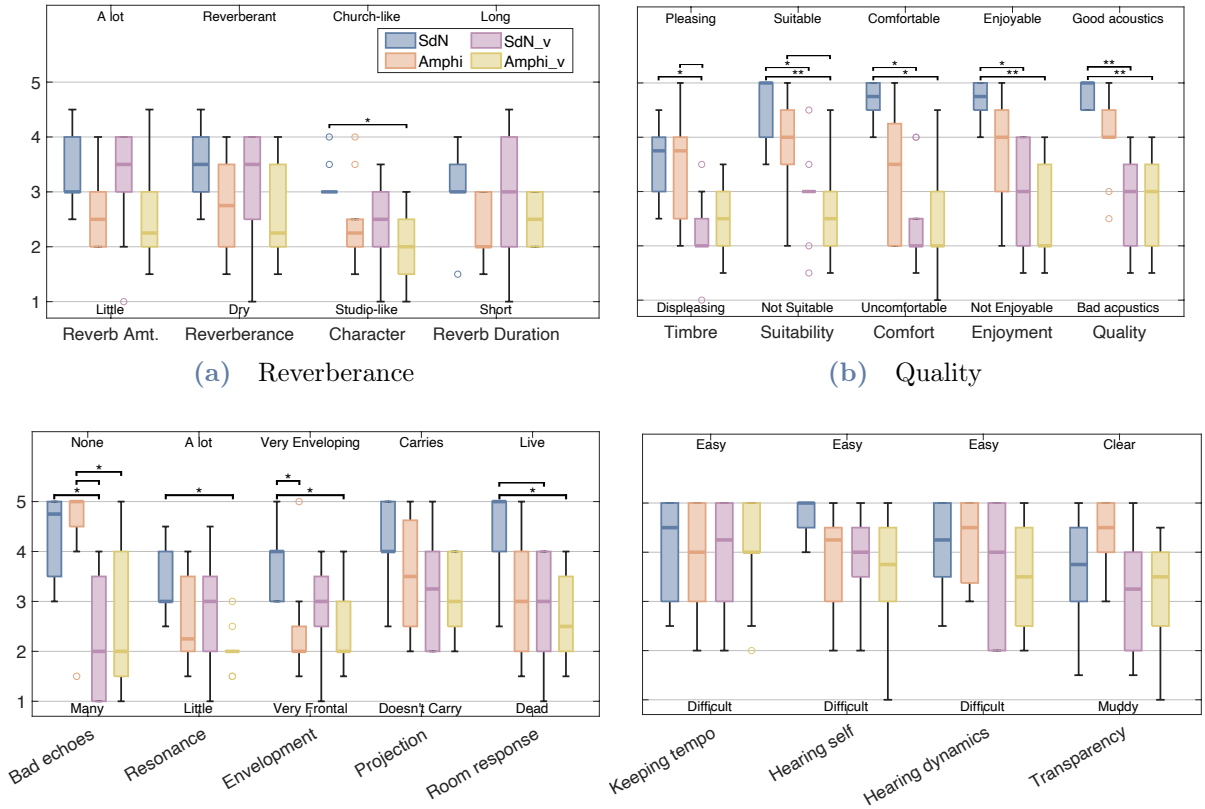
4.1.4 Music performance analysis

Two primary strategies were used to objectively analyze the music performances. The first was to extract a large number of low-level features and train a machine learning classifier to predict which room each performance occurred in. The accuracy of the classifier then served as a proxy measure for the contrast of performances among the rooms which provided a starting point for more in-depth analysis. This strategy is outlined in chapter 5.

The second main approach was to develop a few mid-level features based on musicological principles to shed light on performance characteristics important to historically informed baroque performance. The goal of this strategy was to identify if the room played any effect on the musicians' ability to play in a baroque-appropriate style. This approach is outlined in chapter 6.

4.2 Questionnaire results

The responses to the various questionnaires will be detailed here, followed by a discussion of these results. Each of the questionnaires were designed in english and translated to french with the assistance of native french speakers. They were administered in french to the participants, who were all native or fluent french speakers. The results are reported here in english. The full phrasing of all questionnaires in both french and english can be found in appendix A. Because there were only 10 participants, the results of a statistical significance tests could potentially be



(a) Reverberance **(b)** Quality

(c) Support **(d)** Ease of playing

Figure 4.1 Box plots of responses to the acoustic rating questionnaire (see tables A.2 and A.3) for the four acoustic settings. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima. Brackets indicate a p-value of $< .05$ with asterisks (*) and double asterisks (**) indicating p-values of $< .01$, and $< .001$, respectively according to a Kruskal-Wallis test with Bonferroni correction applied.

misleading. Therefore, the analysis of the questionnaires primarily examines the distribution of data provided by figs. 4.1 to 4.4.

4.2.1 Rating questionnaire results

The acoustic rating questionnaire results are shown in fig. 4.1. The terms are organized into four different groups, which are the result of the factor analysis done by Schärer Kalkandjiev

and Weinzierl (2018). The groups are: reverberance, quality, support, and ease of playing. Significance levels are shown according to a Kruskal-Wallis test with Bonferroni correction.

Within the *reverberance* group (fig. 4.1a), the participants generally rated the amphitheater as being less reverberant than the Salon des Nobles. This corresponds with the measured T and EDT of the two rooms (see fig. 3.4). Additionally, the rated reverberance of the virtual simulations of these two rooms match up fairly well with their real counterparts, although the variability of the virtual responses tends to be a little higher. This is a good indication that the VAEs are at least well-calibrated for this acoustic parameter.

The *quality* group (fig. 4.1b) is perhaps the most telling of the participants' acoustic experience among these four different settings. The Salon des Nobles was rated significantly higher in most categories than one or both virtual environments. The overall quality of the acoustics of the virtual halls was rated lower than the two real rooms. This suggests that, despite achieving a relatively similar level of reverberance to their real counterparts, there were other acoustic properties that the virtual halls did not faithfully reproduce. The ratings in this groups suggest that the participants regarded the acoustics of the Salon des Nobles very highly.

The ratings within the *support* group (fig. 4.1c) reinforce the findings from the acoustic measurements. In general, the Salon des Nobles was rated as significantly more enveloping, more resonant, and as having a livelier room response than one or more of the other environments. The difference in envelopment between the two real halls is particularly of note and may be due to the higher inter-aural cross correlation ($IACC$) and lateral energy fraction (LEF) in the Salon des Nobles (see table 3.2) due to these parameters' associations with *spaciousness*. The more lively room response rating in the Salon des Nobles may be due to the higher late support (ST_{Late}) parameter values (see fig. 3.8b) as this value has been found to be a good indicator of how musicians perceive the hall from their stage position. Another significant finding in this group is within the "bad echoes" category (this is written as "bad echoes" in fig. 4.1c for brevity but was actually written as "echoes/disturbing reflections" in the questionnaire given to the participants; full questionnaires are available in tables A.2 and A.3 in appendix A) where the virtual environments were rated significantly lower than the real environments. This suggest a deficiency with the virtual reproduction for this category.

Among the categories within the *ease of playing* group (fig. 4.1d), there were no significant differences between the ratings of the various acoustic settings. However, there is a tendency for the participants to rate the Salon des Nobles as easier to hear oneself in ("hearing self" in fig. 4.1d for brevity) than the other settings. This indicates musicians had an easier time hearing themselves in the Salon des Nobles compared to the amphitheater and the virtual

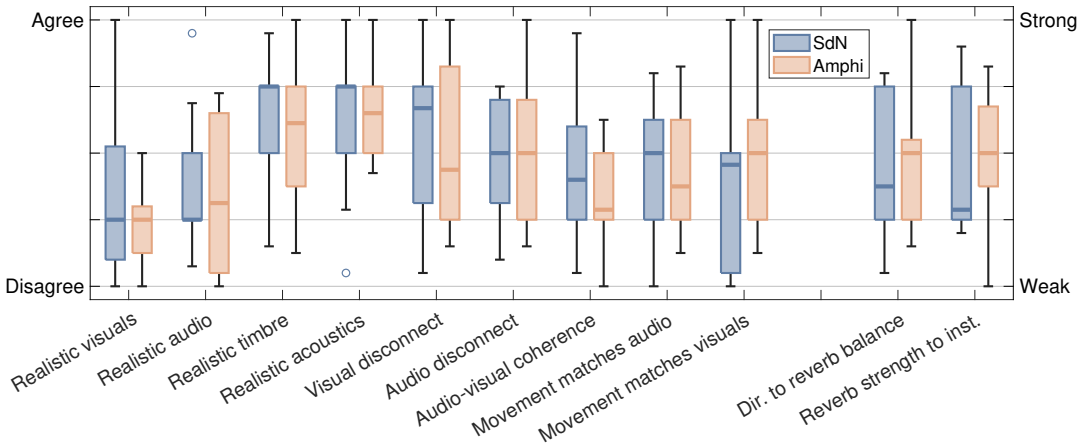


Figure 4.2 Box plots of responses to the virtual questionnaire (see table A.4) for the two virtual environments. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima.

environments. This may be partly due to the higher early support (ST_{Early}) parameter values (see fig. 3.8a) since this value has been found to be a good indicator of how well musicians can hear themselves and each other.

4.2.2 Virtual questionnaire results

The virtual questionnaire was intended to evaluate the virtual environment across a number of categories, including those which are known to be troublesome in virtual environments such as audio and visual latency as well as overall realism. The results of the virtual questionnaire are shown in fig. 4.2. In general, the results indicate a fairly low or mediocre impression of the virtual environments. The worst rated categories are visual realism and audio realism. Despite the low ratings for audio realism, both timbre realism and acoustic realism were rated fairly moderately.

The ratings between the two rooms are relatively similar in all categories. No significant differences were found between ratings in the two virtual environments according to a Wilcoxon signed-rank test. The entire questionnaire containing the full phrasing of each question in english and in french can be seen in table A.4 in appendix A.

Interpretation of these results is difficult on an absolute scale without direct comparisons to other VAEs. In addition, while overall quality was desired, ecological validity (*i.e.* that the

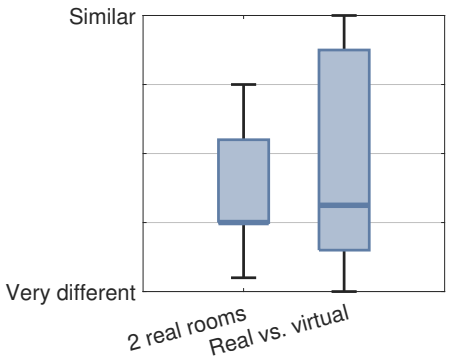


Figure 4.3 Box plots of responses to the comparison questionnaire (see table A.6) asking about the similarity between the two real environments and between the real and virtual environments. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima.

acoustics' effects on performance characteristics would be comparable in the real and virtual environments) was the intended goal of the auralization system.

4.2.3 Comparison questionnaire results

The comparison questionnaire, which was given to all participants at the end of their last session, was intended to evaluate the perceptual similarity of the various environments. The participants were asked to compare the similarity of the two real rooms as well the real rooms and their virtual counterparts. They were asked to rate these on a 5-point scale ranging from "very different" (1) to "very similar" (5).

The results (shown in fig. 4.3) indicate that the two real rooms were rated as slightly dissimilar (a mean value of 2.45). The responses showing the similarity between the real rooms and their virtual counterparts show a fairly broad range of ratings. Some participants rated them as being fairly different, with half of the participants giving a rating of 2 or lower, while some participants rated them as being quite similar, with four participants giving a rating of 4 or higher.

Some of these high ratings are perplexing since they contradict other comments made by the same participants. For example, one participant who gave a rating of 5/5, indicating the real environments were similar to their virtual counterparts, also described the virtual environments as having an "overall metallic sound that lacks roundness." Another participant who rated the

real and virtual environments as being similar (4.6/5) said about the virtual environment that, "if it's for professional work, it's not quite there yet." And finally, a third participant who gave a rating of 4/5 on the similarity scale said they felt that certain aspects of the virtual environment were "a bit disturbing." These comments do not indicate the participants felt that the virtual acoustics were a faithful rendering of the real acoustics, and yet their ratings indicated that.

On reevaluating the questions that were asked, it seems likely that, due to the phrasing of the question, some participants misunderstood the question. The original question, was proposed as: "How do you assess the difference between your playing experiences (musical performance) in both contexts (real rooms versus virtual clones)." In future experiments, the phrasing of questions should be given more consideration.

4.2.4 Presence questionnaire results

The results of the ITC-SOPI questionnaire are broken down into four factors which make up a sense of presence or of "being there." These factors are: *sense of physical space*, *engagement*, *ecological validity*, and *negative effects*. Each of these are based on ratings from 1 to 5 with higher values corresponding to a better sense of presence. These questions were not referring specifically to the audio component of the system, but rather the combined effects of the audio and visual components on the overall sense of presence.

These results (seen in fig. 4.4) further suggest a somewhat inferior impression of the virtual environment by the participants. Three of the four categories have an average of around 2 out of 5 while the remaining, best-performing category, *engagement* is closer to (but still below) 3.

4.3 Virtual environment evaluation

The combined results of the ratings questionnaires (particularly figs. 4.1b and 4.4) and the responses to the open-ended questions (reported in tables 4.4 and 4.5) indicate an overall impression of the virtual environment that is somewhat mediocre. These responses have been helpful in identifying areas where improvement is needed. This section is devoted to a cursory evaluation of these responses.

One of the clearest indicators that the virtual environment was flawed is in the responses of the acoustic rating questionnaire, particularly within the categories under the quality group (fig. 4.1b). In practically every category in this group the virtual environments received significantly lower ratings overall than the real halls. Furthermore, the responses to the virtual

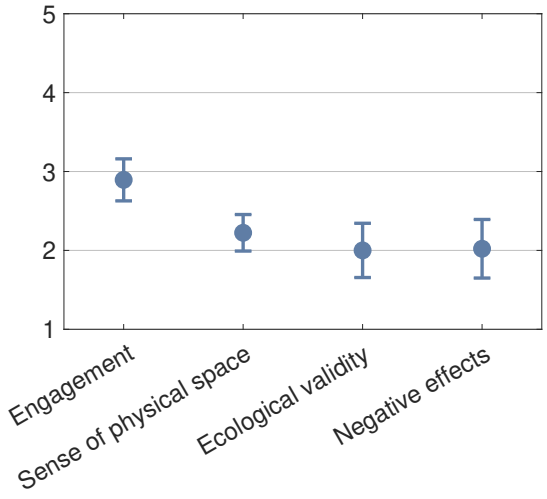


Figure 4.4 Means and 95% confidence intervals of responses to the presence questionnaire.

questionnaire (fig. 4.2) and the presence questionnaire (fig. 4.4) suggest a fairly mediocre impression of the virtual environment.

In response to the open-ended question which asked about the impact of the virtual environment on their performance, the responses of which can be found in table 4.4, several participants described an effort to intentionally ignore the virtual environment as a coping mechanism since they found it distracting. For example, flutist 2 said they played as if they were in a studio, "trying to ignore the sound I was hearing." Elaborating further, this same flutist said, "the principal effort was not to be influenced by what one saw or heard." Theorbist 1 stated that they had to exert "a greater concentration to stay in my bubble and not participate in the environment." This indicates that the virtual environment had a profound negative effect on the ability of some of the participants to perform as if in a real environment. This also signifies a failure of the system to provide a reasonable alternative to studying musicians' interactions with acoustics as in a real space.

One major issue with the auralization system that many participants complained of was the unnatural spectral coloration of the virtual acoustics. This was also an issue in the preliminary study where all four participants described the virtual acoustics as sharp or metallic (see section 3.8). In this primary study, five out of ten participants (two violists, two flutists, and one theorbist) complained about a high-frequency coloration of the auralization using words such as "sharp," "metallic," and "sibilant." These comments were present in questionnaires of both of the virtual spaces.

Table 4.4 All participant responses to the question, "Has being in a virtual environment had an impact on your performance? If so, how?"

Participant	Salon des Nobles	Amphitheater
Violist 1	More concentration perhaps? Imagination is preferred.	I found this performance rather complicated, because it interfered with my imagination.
Violist 2	A little, mostly the fact that I was hearing more from the right.	Fairly little.
Violist 3	Not really. Being often in a studio with headphones and sometimes even with an added acoustic. I'm therefore used to it.	A little seasickness = adjustments to the visuals to refine.
Violist 4	I had a lot of difficulty really projecting myself into the environment, with the impression of being half in two places, it became difficult to concentrate on my playing.	The virtual environment was hard to concentrate in but it turned out to be more fun in this modern configuration. I was more free than in the previous one.
Flutist 1	Same as in the first pass, I didn't succeed on projecting visually or projecting the sound like a real room would allow. Additionally, I didn't really notice differences between the two presented acoustics.	I didn't really succeed in projecting in this room neither visually or through sound. It's more the visual side that posed a problem than the sound, the sound rendering was pleasant but did not allow me to project.
Flutist 2	If this was for a game, it would be fun. If this was for professional work, it is not quite there yet. The principal effort was to not be influenced by what one saw or heard.	I played like I was in a studio, trying to ignore the sound I was hearing.
Flutist 3	Difficult for the imagination. The virtual acoustics are rather destabilizing at first.	The effect of the hissing in the reverberation is quite unnatural.
Theorbist 1	Maybe a greater concentration to stay in my bubble and not participate in the environment. Comparable to a solo recording session.	Playing in front of real people changes EVERYTHING. Music then becomes really a language that speaks to them.
Theorbist 2	No.	No.
Theorbist 3	(No answer.)	Yes, lack of concentration due to the distracting noises but also a taste for the playful side to observe and analyze what happens while playing.

Nine out of ten participants had complaints about the sound in less precise terms, citing such issues as "intrusive sounds" and "frequencies that are resonating too long." While these complaints may not be related to the coloration of the system, this should not be completely ruled out as non-expert listeners may not be equipped with the appropriate vocabulary or experience to describe the problem accurately. Furthermore, no participants reported a coloration that was contrary to this, in other words, no participants reported a coloration that was dull or that contained too much low-frequency energy.

An investigation into the coloration after the preliminary experiment suggested that the problem was with a poorly calibrated feedback delay network (FDN) (see section 3.8). However, this could not be the case in these two virtual environments since no FDN was used. This suggests that there is another significant source of coloration. However, a full investigation into the source of this coloration was out of the scope of this thesis.

4.3.1 Suggestions for improvements

One difficulty in assessing the success of a virtual environment such as the one used in this thesis, is that the level of realism needed to perform these studies such that they would produce reliable results is not yet known. The gap between plausibility and perfection is large, and the necessary level of realism for these types of studies lies somewhere in that gap. Whereas numerous perceptual studies have been carried out using auralizations, most of these tend to be performed from the point of view of a passive observer or audience member rather than an active participant and sound source. More studies on the limits of perception of various aspects of VAEs from the point-of-view of an active participant would be valuable in designing and developing future such systems.

When asked about how the VAE could be improved, participant responses suggested that the spectral coloration was the primary problem. Additionally, three other areas of improvement were identified by the participants. The first was the sound of the video projector, which, despite an absorbing baffle, some participants complained was too loud and distracting. The second was the poor quality of the visuals, which several participants critiqued and one participant suggested eliminating entirely. The third was that the virtual acoustics of the two rooms were not distinct enough, and that several of the participants did not notice much of a difference between them. This last point suggests that even beyond the coloration issue, there may be additional refinements of the system required to render the virtual acoustics more faithfully. Indeed, the responses to the acoustic rating questionnaire (see fig. 4.1) show that the ratings for the virtual environments are rarely significantly different from each other, while the two

Table 4.5 All participant responses to the question, "Do you think the virtual acoustic environment can be improved? If so, how?"

Participant	Salon des Nobles	Amphitheater
Violist 1	Round out the sharpness.	Less strength in the sharpness and more resonance.
Violist 2	(No answer.)	(No answer.)
Violist 3	It is necessary to know the acoustics of the original space.	A little "I don't know what" to improve.
Violist 4	Yes, by working on the intrusive sounds that are very disruptive while playing, and more precision in the rendering of the sound, notably in the resonance and reverberation.	Yes, still because of the resonating echoes that are sharp and metallic that distract from the performance. The rest, in the reverberation being more realistic, more reactive.
Flutist 1	Even if the reverberation is pleasant, it could be more faithful to the real rooms because the two proposed acoustics were very similar. Perhaps also more virtual realities for an easier sound and visual projection.	The acoustic environment could be improved visually for a more realistic rendering and the suppression of intrusive sounds.
Flutist 2	Get rid of intrusive noises (fan, microphone too close too the mouth), eliminate the frequencies that are resonating for too long.	Probably first by improving the capture of the sound at the mouth ... putting the microphones at the level of the musician's ears? Highlighting nearby noises would be a priority.
Flutist 3	Perhaps the virtual acoustics could benefit from resembling more the real acoustics.	The effect of the hissing in the reverberation is quite unnatural.
Theorbist 1	By getting rid of the machine sounds.	By getting rid of the image.
Theorbist 2	Yes, the changes in acoustics that were not very different and, to the contrary, I felt a little bothered by them.	By eliminating noises not normally heard.
Theorbist 3	Yes, lots of intrusive noises and increased high harmonics according to my perception, more lower harmonics would be desirable. Quality of the sound is quite rough and missing a little "roundness."	Less intrusive sounds and sharp harmonics? Sound overall "metallic" lacking in "roundness."

real environments do show different ratings in a number of categories.

The sound of the projector was noted by several participants as quite loud and distracting. This is a factor that should be taken into consideration when contemplating the inclusion of visuals in a VAE. The resolution of the projector was 1080p, which, when projected onto a screen over 3 m wide, results in a rendering in which individual pixels were easily seen. Considering that the quality of the visuals was rated poorly, and the number of comments which cited it as an unwelcome distraction, the inclusion of a visual rendering was probably more damaging than helpful. In order to justify the inclusion of visuals in future experiments, the quality of the visual rendering must be taken into consideration with other potential advantages and disadvantages (see section 3.6.1).

One of the novel components of the experimental virtual archaeological-acoustics (EVAA) system was the implementation of dynamic directivity. However, the effect of this was difficult to assess in this study simply because the musicians generally remained stationary during their performances. A better scenario, in which more movement is natural and expected, should be used to better evaluate the effect of dynamic directivity in a future experiment.

4.4 Discussion

While the participants generally rated the virtual environments poorly, their feedback was valuable in identifying problems, a necessary step in order to make future improvements. Furthermore, the experiences in the real rooms still appear to be valuable for further analysis.

One major finding from the experiment thus far is that the participants seem to have found the acoustics of the Salon des Nobles as more appropriate to their playing than the acoustics of the amphitheater in Cité de la Musique. This is evident in the responses to the acoustics rating questionnaire (see fig. 4.1) and in the responses to the open-ended questions which are reported in full in section 5.4. It should be noted, however, that these ratings were not given blind and that the results could certainly be influenced by visual elements and the historical nature of the Château de Versailles.

While serious issues with the auralization system were identified during this experiment, a full investigation with proposed solutions and follow-up evaluations was beyond the scope of this thesis, for which the primary goal was to investigate the effect of acoustics on musical performance. Unfortunately, the negative ratings of the virtual environments indicate that musicians were unable to concentrate on their performance to a degree that would allow for fruitful analysis. As a result, the music performance analysis in chapters 5 and 6 will exclusively

focus on recordings from the two real rooms.

Chapter 5

Music performance analysis

This chapter describes the first approach taken towards analyzing the music performances recorded in chapter 4. In this approach, a number of mostly low-level features were extracted from the audio which were then subject to dimensionality reduction. The resulting dimensions were then used to train a supervised learning model to predict which room the performance occurred in. The success of these models served as an indicator of the difference in performance style between the two rooms. Additionally, a subset of the original features was subjected to statistical analysis and analyzed in the context of participant comments.

Performing any statistical analysis, including training machine learning classifiers, directly on time domain audio signals would not yield satisfactory results as these signals are noisy and do not contain enough discriminative information. Instead, a number of discriminating features were first extracted directly from the raw audio files of the recordings of the performances (see section 5.1). These features were then subject to dimensionality reduction, both through manual selection and more advanced techniques such as principal component analysis (PCA) (see section 5.2). These dimensionally-reduced feature sets were then used to train support vector machine (SVM) classifiers tasked with predicting the room the performances occurred in. The results served as a proxy measure for how different the performances were from one room to another and provided a path to more in-depth analysis. The performances of each musician were also examined individually and those results are reported in section 5.4.

Table 5.1 List of features extracted from the recordings of performances used in the music performance analysis.

Timbre		Tempo	Intensity
Zero-crossing rate	Spectral rolloff	Note-level tempo	RMS
Spectral brightness	Spectral centroid	Inter-onset-interval (IOI)	
Spectral spread	Spectral skewness	Normalized IOI	
Spectral kurtosis	Spectral flatness		
MFCCs 0-20			

5.1 Feature extraction

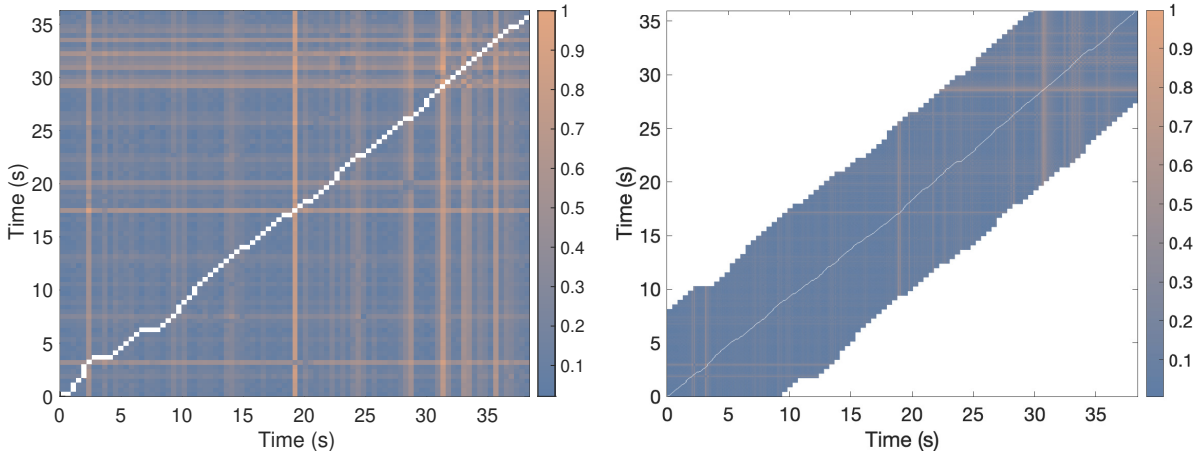
Carl Seashore's pioneering work in music performance analysis often divided a musical signal into the constituent parts of pitch, loudness, time, and timbre (Seashore, 1938). The extracted features (listed in Table 5.1) can broadly be grouped to represent three of these four categories: loudness, time, and timbre. Although the terms tempo and intensity are preferred to time and loudness, as they are more precise. The remaining attribute of pitch was thought to be less important in this instance since the musicians in this study relied on a musical score from which significant deviations of pitch were not expected. However, some pitch analysis was performed in chapter 6, namely in the form of vibrato analysis.

5.1.1 Synchronization

In order to calculate features related to tempo, it is necessary to know the precise timing of each onset of interest, such as the downbeat of each measure or the start of each note. This can be facilitated by synchronizing the audio recording of the performance and the musical score. Aside from making tempo feature calculation easier, score synchronization also makes possible additional types of analysis, including dimensionality reduction and easier comparison of features between performances.

In order to have a digital representation of the score, a musical instrument digital interface (MIDI) file for each piece was manually created using the notation software Sibelius[1]. Note onset times and note values were extracted from the MIDI files in MATLAB using the MIDI toolbox developed by Eerola and Toiviainen (2004a,b). The digital score was processed to remove any trills or ornaments, so that each onset would have a precise note value. Audio files

[1] https://www.avid.com/sibelius



(a) Reduced resolution similarity matrix with DTW path.

(b) Full resolution similarity matrix on reduced neighborhood with DTW path.

Figure 5.1 Example of of multi-scale DTW.

were created from the unprocessed MIDI files (so as to maintain trills and ornaments) using software synthesizers that resembled the appropriate instrument, turning the audio-to-score synchronization problem into an audio-to-audio synchronization problem for which there are many strategies.

The synchronization was performed using an audio-to-audio matching algorithm developed by Dixon and Widmer (2005), custom-coded in MATLAB. It uses a first-order difference of a non-linearly warped discrete Fourier transform (DFT) to create a representation of the audio file. This representation is then used to calculate a similarity matrix between the two audio files. A dynamic time warping (DTW) algorithm is then used to find the least cost path in this matrix which provides a synchronization between the timing events in both audio files.

The calculation time for such matching algorithms can be significant due to its high computational cost. To ameliorate this, a version of multi-scale DTW (also known as fast DTW) was used which can speed up calculation time by a factor of 40 to 100 without a loss of accuracy (Wang, Ewert, et al., 2016). Multi-scale DTW first projects a path onto a coarser resolution version of the similarity matrix (see fig. 5.1a). This coarse resolution path is then used to create a neighborhood on the full similarity matrix to which the search for the least cost path is constrained (see fig. 5.1b).

In an evaluation of the accuracy of different audio-to-audio alignment methods, the method from Dixon and Widmer (2005) was shown to be one of the most robust and accurate (Kirchhoff

and Lerch, 2011). While more precise methods were found in that study, they were largely based on pitch features, such as pitch chroma, which must be calibrated against a standard pitch (such as A440) and tuning systems. Since historically informed performances of baroque music typically use a different tuning standard, commonly A415 (Haynes, 2007, p. 44), and different tuning systems from most modern music, the matching algorithm from Dixon and Widmer (2005) was chosen as it avoids this pitch calibration problem while also maintaining relatively high accuracy and robustness. Since the matching algorithm is not perfectly accurate, the detected note onsets were manually verified and adjusted, where necessary, using Sonic Visualizer [2].

5.1.2 Tempo

Rather than a global measure of the average tempo of the whole piece, which would fail to capture expressive nuances, the **note-level tempo**, measured in beats per minute (BPM) was used as defined by,

$$BPM_{note}(i) = \frac{60s}{t_o(i+1) - t_o(i)} * \Delta\tau_{i,i+1} \tag{5.1}$$

where $t_o(i)$ is the onset time of the score event of interest (in this case, each note) and $\Delta\tau$ is the distance between two score events, measured in beats. The specific beat unit could change depending on the time signature and the nature of the piece of music. For example, in a 4/4 piece of music where the tactus, or most salient periodicity, is a quarter note, a quarter note would have a value of 1 while an eighth note would have a value of 0.5 and a half note would have a value of 2.

Additionally, the **inter-onset-interval (IOI)**, which measures the distance, in seconds between two onsets of interest, in this case notes, were calculated:

$$IOI(i) = t_o(i+1) - t_o(i). \tag{5.2}$$

Lastly, the **normalized inter-onset interval**, which normalizes each onset by the distance between onsets in beats, was used. It is defined as:

$$IOI_{norm}(i) = \frac{t_o(i+1) - t_o(i)}{\Delta\tau_{i,i+1}}. \tag{5.3}$$

Although these tempo features are likely highly correlated, both absolute and normalized

[2] https://www.sonicvisualiser.org/

representations of tempo have shown value in previous studies (Timmers, 2005).

5.1.3 Timbre

In order to characterize the timbre of the different recordings, a number of features which are commonly used in music information retrieval (MIR) were enlisted. These features generally describe the distribution of energy of the magnitude spectrum. They were extracted using the MIRToolbox (v.1.8.1) [3]. A summary of these features follows, while more detailed explanations can be found in Lerch (2012).

Zero-crossing rate is the number of times the signal changes sign in a given window. A higher value implies more high frequency content in the signal while the variability of the zero-crossing rate is an indication of its periodicity. The zero-crossing rate is calculated thus:

$$ZCR(n) = \frac{1}{2 \cdot K} \sum_{i=i_s(n)}^{i_e(n)} |\text{sign}[x(i)] - \text{sign}[x(i-1)]| \tag{5.4}$$

where x is an audio signal, n is the frame, i is the sample within each frame, K is the block size, and the sign function is defined by

$$\text{sign}\ [x(k)] = \begin{cases} 1, & \text{if } x(i) > 0 \\ 0, & \text{if } x(i) = 0 \\ -1, & \text{if } x(i) < 0 \end{cases} \tag{5.5}$$

Spectral rolloff serves as a measure of the timbral bandwidth of a signal. It is technically defined as the frequency bin below which the cumulative sum of magnitudes of the short-time fourier transform (STFT) of the signal reach a percentage (typically 85 %) of the total sum of STFT magnitudes. It is calculated as:

$$SR(n) = i \Big|_{\sum_{k=0}^{i} |X(k,n)| = 0.85 * \sum_{k=0}^{K/2-1} |X(k,n)|} \tag{5.6}$$

where k is the frequency bin. It is commonly expressed in Hz where lower values indicate a smaller bandwidth.

The **spectral centroid** is often described as the "center of gravity" of the spectral energy of a signal and is correlated with the perceived brightness of a signal:

[3] https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox

$$SC(n) = \frac{\sum_{k=0}^{K/2-1} k \cdot |X(k,n)|^2}{\sum_{k=1}^{K/2-1} |X(k,n)^2}. \tag{5.7}$$

Based on the spectral centroid is the **spectral spread** which measures the concentration of the power spectrum surrounding the spectral centroid:

$$SS(n) = \sqrt{\frac{\sum_{k=0}^{K/2-1} (k - SC(n))^2 * |X(k,n)|^2}{\sum_{k=0}^{K/2-1} |X(k,n)|^2}}. \tag{5.8}$$

Spectral skewness is a measure of the symmetry of spectral energy, calculated as

$$SSk(n) = \frac{2\sum_{k=0}^{K/2-1} (|X(k,n)| - \mu_{|X|})^3}{K \cdot \sigma_{|X|}^3}, \tag{5.9}$$

where μ is the arithmetic mean and σ is the standard deviation.

Spectral kurtosis is a measure of the gaussianity of the distribution of spectral energy, defined by

$$SK(n) = \frac{2\sum_{k=0}^{K/2-1} (|X(k,n)| - \mu_{|X|})^4}{K \cdot \sigma_{|X|}^4} - 3. \tag{5.10}$$

Spectral flatness is a description of how noisy or tonal a signal is, where a higher value indicates a noisier signal. It is formally the ratio of the geometric mean and arithmetic mean of the magnitude spectrum, defined by:

$$SF(n) = \frac{\sqrt[K/2]{\Pi_{k=0}^{K/2-1} |X(k,n|}}{2/K \cdot \sum_{k=0}^{K/2-1} |X(k,n)|} = \frac{\exp\left(2/K \cdot \sum_{k=0}^{K/2-1} \log(|X(k,n)|)\right)}{2/K \cdot \sum_{k=0}^{K/2-1} |X(k,n)|}. \tag{5.11}$$

A set of timbre features with a stronger perceptual basis, called **mel-frequency cepstral coefficients (MFCCs)**, were also used. MFCCs are similar to the cepstrum in that they are a logarithmic representation of the spectrum, however, they are based on a nonlinearly warped frequency spectrum derived from the mel scale. The mel scale is an empirically obtained frequency spacing judged by listeners to be the same distance apart and therefore has a very strong perceptual connection (O'Shaughnessy, 1987). MFCCs are calculated by first computing the mel spectrum with a group of overlapping triangular filterbanks then taking the logarithm of the magnitude of each frequency band. Finally, a discrete cosine transform (DCT) is applied to each of these resulting frequency bands. Implementation of MFCCs can vary due to a number

of factors including the number of filters, the start and end frequencies, and the filterbank normalization, but are generally calculated for coefficient j as,

$$MFCC_j(n) = \sum_{k'=1}^{K'} \log(|X'(k', n)|) \cdot \cos\left(j \cdot \left(k' - \frac{1}{2}\right) \frac{\pi}{K'}\right) \tag{5.12}$$

where $|X'(k', n)|$ is the mel-frequency warped magnitude spectrum. While these features were originally developed for use in speech processing, they were first applied to music-related problems in the early 21$^{\text{st}}$ century (Logan, 2000) and their success and popularity in modeling musical properties has only grown since then (Siedenburg et al., 2016). This analysis made use of 21 coefficients, including the 0$^{\text{th}}$ coefficient which is more strongly correlated with the total energy. A relatively high number of coefficients was chosen as the intention was to run these features through PCA. A higher number of features would be somewhat desirable in that case, so as to make use of any available variance in the data without a high risk of overfitting, which is a typical drawback of using too many features.

5.1.4 Intensity

Intensity features are commonly used in music analysis since they are directly related to the musical parameter of dynamics. One of the most commonly used intensity feature is root mean square (RMS) (Lerch, 2012, p. 73) and is calculated as follows:

$$RMS(n) = \sqrt{1/K \sum_{i=i_s(n)}^{i_e(n)} x(i)^2}. \tag{5.13}$$

Other models of intensity exist including a digital implementation of a voltage unit (VU) meter (implemented following Lobdell and Allen (2007)), a psychoacoustic model of loudness proposed by Zwicker and Fastl (1999) and standardized in ISO 532-1:2017 (ISO, 2017), and a loudness measure described in an International Telecommunication Union (ITU) recommendation (ITU-R, 1998), however, in previous iterations of this analysis, these models were found to be strongly correlated with RMS. For the sake of simplification, the number of features in the final analysis was reduced to use only the RMS measure to describe intensity.

5.2 Methodology

Prior to extracting features, each audio file was edited to remove silence before and after each performance. Due to an oversight in the experiment planning process, the microphone recording level was not calibrated, so any measured differences in overall level may be due to the different gains used across sessions or a slight deviation in distance between the instrument and the microphone, rather than the musician actually playing at a different volume. In order to account for this, the audio files were normalized to have the same overall RMS level as a way to reduce inter-file intensity differences while maintaining the changes in intensity within each recording. This is not an ideal solution as it reduces the actual differences in overall intensity from one performance to another, however, it was seen as a reasonable solution to the lack of microphone level calibration. Future experiments would be advised to carefully calibrate the input recording level so as to have more dependable data on the overall intensity of the signals recorded.

The timbre and intensity features were extracted from audio frames with a window size of 2048 samples and hop size of 128 samples. A Hamming window was applied prior to transformation into the frequency domain. Rather than applying analysis directly to these small frames, which may not carry much useful information, frames were combined into lengths of musical durations. These durations were either of 16^{th} or 32^{nd}-notes, depending on the composition and its tempo (faster pieces used 16^{th}-note durations while slower pieces used 32^{nd}-note durations). Combining smaller frames into larger, more meaningful segments has been shown to improve music classification tasks, where these segments are often referred to as "texture windows" (Tzanetakis and Cook, 2002; Scaringella and Zoia, 2005). Using the note onset information, which was used to calculate tempo (see section 5.1.2), all feature vectors (including the tempo features) were transformed to this new time grid via linear interpolation.

Aside from providing a more meaningful segment upon which to apply further analysis, this process also synchronizes all feature vectors of a single composition, making them the same length. This makes direct comparisons easier and also facilitates further processing, such as dimensionality reduction techniques.

Overall, 33 features were extracted from the audio files (see table 5.1). While these features were selected based on their common usage in the field of MIR, it was not known beforehand which features would be particularly discriminative for the task at hand. Furthermore, it is likely that some features may be strongly correlated and therefore redundant. In order to reduce the number of features while still maintaining most of the variance of the data, PCA

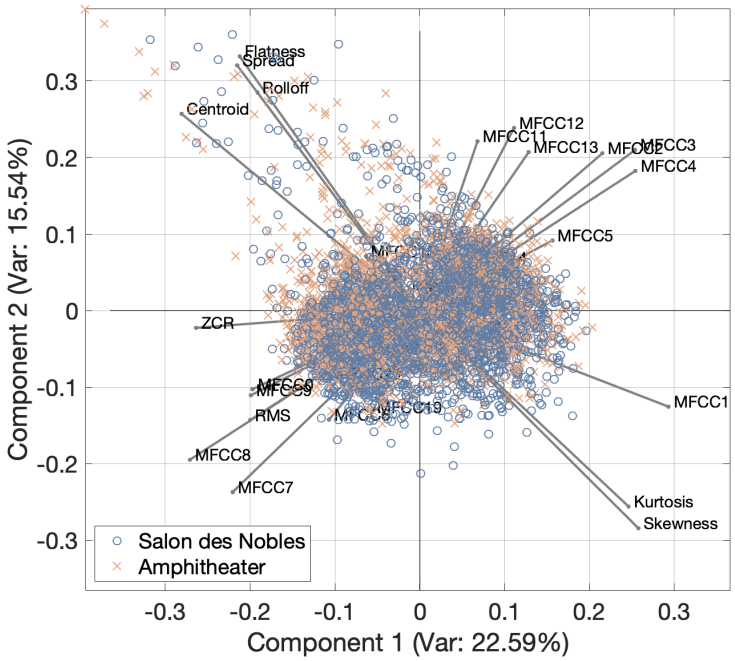


Figure 5.2 A biplot of the top two principal components from a PCA performed on all flutists' performances of Animé.

was performed. A subset of the resulting components (all of those with an eigenvalue greater than 1) were used as the new feature set. Figure 5.2 shows an example of the top two principal components from a PCA performed on all flutists' performances of Animé. For each observation in the dataset, the performance space is indicated. In this example, the top two components are responsible for 38.13% of the overall variance. Eight components had an eigenvalue > 1; these components were responsible for a combined 79.85% of the overall variance of the data.

Both unsupervised learning (i.e. clustering) and supervised learning (i.e. classification) have been applied towards identifying similarity in music. Supervised learning is particularly useful when the classes are known beforehand, such as in some genre classification tasks (Tzanetakis and Cook, 2002; Burred and Lerch, 2003). Following these same principles, a supervised learning approach was chosen since the class of interest (the different acoustic setting) is already known. Therefore, the feature set was used to train SVMs to predict which room the performance occurred in. The output could then serve as a proxy for how different the performances are among the two different acoustics and help identify areas of interest for more

in-depth analysis.

The data was first partitioned into three groups, one for each instrument. These classes were treated separately in the following analysis. Within each instrument class, the classifiers were trained on four different subsets of the data:

- All compositions and all musicians
- All compositions and individual musicians
- Individual compositions and all musicians
- Individual compositions and individual musicians

Each of these subsets were subject to PCA to reduce the number of features and this reduced feature set was used to train the classifiers, as previous work has shown that performance adaptations to acoustics can be highly individual (see section 2.3).

SVMs are supervised learning models capable of performing linear and nonlinear binary classification. The SVMs in this study utilized radial basis function (RBF) kernels. The classifiers were all trained using 10-fold cross validation in which 90% of the data is selected randomly to be used as training data while the remaining 10% is used as testing data. This is repeated 10 times until all of the data have been used as either testing or training data. The features of each texture window were classified individually and a majority class of all texture windows within a single performance were used to decide the class of that performance as done in previous work in music classification (Tzanetakis and Cook, 2002; Scaringella and Zoia, 2005; Burred and Lerch, 2003).

Any accuracy above random chance (50% in this case) would indicate that the information capture by the features is measuring differences as a function of the room. Further analysis would be needed to understand what these difference are and whether they are factors of interest (i.e. changes in performance style) or not.

The full feature set contains many timbre related features which may be influenced by the sound of the room incidentally picked up by the microphone used to record the instrument. Although the microphone selection and position were chosen to provide a strong direct-to-reverberant ratio, the sound of the room is still somewhat evident in the recordings. It is possible the timbre features may be influenced by this, resulting in classification accuracies that are misleadingly high and not strongly based on factors of interest (i.e. performance characteristics). Therefore, in addition to the full feature set (reduced by PCA), two other feature sets were tested which did not include any timbre features.

Two additional feature sets were used in the analysis: one with intensity and tempo features, and the other with only tempo features. Since it is possible that the intensity features could be influenced by the sound of the room as well, the analysis was also restricted to the tempo-only feature set, which would be completely free of influence from room sound in the recordings. The influence of this room sound on the classifiers is complex and difficult to predict and account for in the ensuing analysis, however, a more in-depth investigation was carried out to better understand this issue (see section 6.5).

5.3 Overall results

The overall accuracy (the number of correct classifications divided by the total number of classifications) of the classifiers trained on the full feature set can be seen in table 5.2. This feature set resulted in almost perfect classification for all viol and flute performances. The results from the theorbo are somewhat mixed but generally fairly high, with a few perfectly classified sets of performances.

The very high accuracy of the classifiers is likely at least partly due to the sound of the rooms' acoustics present in the recordings rather than due to a major and consistent difference among the performances in the two rooms. However, even though these results were possibly biased by factors that were not of interest, there are still conclusions that can be drawn from them.

For example, the relatively lower accuracy of the theorbo performances indicates that the classifiers were not solely relying on the room sound present in the recordings. Although it is also possible that the theorbo didn't excite the room in such a way to manifest a distinction in the room response which was recognizable to the SVM classifiers. The classifiers were almost certainly able to identify some meaningful and consistent differences in playing style among the viol and flute performances (and some of the theorbo performances) in order to achieve such high accuracy. Furthermore, the lowest performing classification tasks among the theorbo performances (around 57%, only slightly above random chance, or 50%) may suggest a lower bound for how much the classifier is influenced by the sound of the room. However, since each instrument excites the room in a slightly different way, both spectrally and spatially, it is possible this lower limit may be different for the other instruments.

One reason for the relatively low accuracy among the theorbo performances may be that the theorbists generally had difficulty playing the chosen repertoire without mistakes. The musicians were therefore less consistent from one repetition to another, resulting in greater

Table 5.2 Results of SVM classifiers for each data subset using full feature set in predicting the acoustic setting of the performance (Salon des Nobles or Amphitheater). The classifier scheme used a majority vote threshold where if the majority of observations in one performance are of the correct class then that performance is counted as a correct classification.

Musician	Viol		Flute		Theorbo	
1st	Couplet 12	100.0%	Animé	100.0%	Gigue A	76.9%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	57.1%
	Couplet 18	100.0%	Gravement	100.0%	Prelude	85.7%
	All	100.0%	All	100.0%	All	92.6%
2nd	Couplet 12	100.0%	Animé	100.0%	Gigue A	57.1%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	100.0%
	Couplet 18	100.0%	Gravement	100.0%	Prelude	100.0%
	All	100.0%	All	100.0%	All	94.7%
3rd	Couplet 12	100.0%	Animé	100.0%	Gigue A	100.0%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	100.0%
	Couplet 18	100.0%	Gravement	100.0%	Prelude	100.0%
	All	100.0%	All	100.0%	All	100.0%
4th	Couplet 12	100.0%				
	Couplet 13	100.0%				
	Couplet 18	100.0%				
	All	100.0%				
All	Couplet 12	100.0%	Animé	100.0%	Gigue A	59.3%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	73.7%
	Couplet 18	100.0%	Gravement	100.0%	Prelude	100.0%
	All	**100.0%**	**All**	**100.0%**	**All**	**95.7%**

variability overall; they were likely concentrating primarily on the technical demands of the music rather than on optimizing their interpretation for the environment.

The classification accuracy of the SVMs trained only on the tempo and intensity features (see table 5.3) is significantly lower than those trained on the full feature set. Overall, the results are still relatively high, suggesting the presence of some musically meaningful differences in tempo and intensity among the performances in the two rooms. The results for the individual musicians tend to be higher than for those of all musicians combined. This suggests that there

Table 5.3 Results of SVM classifiers for each data subset using only tempo and intensity features in predicting the acoustic setting of the performance (Salon des Nobles or Amphitheater). The classifier scheme used a majority vote threshold where if the majority of observations in one performance are of the correct class then that performance is counted as a correct classification.

Musician	Viol		Flute		Theorbo	
1st	Couplet 12	66.7%	Animé	100.0%	Gigue A	92.3%
	Couplet 13	66.7%	Lentement	100.0%	Gigue B	85.7%
	Couplet 18	83.3%	Gravement	66.7%	Prelude	71.4%
	All	66.7%	All	88.9%	All	85.2%
2nd	Couplet 12	83.3%	Animé	85.7%	Gigue A	85.7%
	Couplet 13	83.3%	Lentement	85.7%	Gigue B	66.7%
	Couplet 18	83.3%	Gravement	71.4%	Prelude	66.7%
	All	83.3%	All	61.9%	All	73.7%
3rd	Couplet 12	83.3%	Animé	66.7%	Gigue A	91.7%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	83.3%
	Couplet 18	83.3%	Gravement	100.0%	Prelude	83.3%
	All	94.4%	All	77.8%	All	70.8%
4th	Couplet 12	100.0%				
	Couplet 13	83.3%				
	Couplet 18	66.7%				
	All	66.%				
All	Couplet 12	70.8%	Animé	63.1%	Gigue A	81.3%
	Couplet 13	70.8%	Lentement	73.7%	Gigue B	89.5%
	Couplet 18	70.8%	Gravement	79.0%	Prelude	63.2%
	All	**66.7%**	**All**	**66.7%**	**All**	**67.1%**

are some individual adaptation strategies that are not common across performers, as found in previous research (see section 2.3). One unexpected outcome of these classification tasks is that the accuracy of some theorbo performances is actually better with this reduced feature set (59.3% for all musicians playing Gigue A, for example) compared to the full feature set that included all timbre features (81.3% for all musicians playing Gigue A).

It is possible that some of the classifiers' performance may be due to the intensity feature being influenced by the sound of the room present in the recordings. For example, a stronger

Table 5.4 Results of SVM classifiers for each data subset using only tempo features in predicting the acoustic setting of the performance (Salon des Nobles or Amphitheater). The classifier scheme used a majority vote threshold where if the majority of observations in one performance are of the correct class then that performance is counted as a correct classification.

Musician	Viol		Flute		Theorbo	
1st	Couplet 12	66.7%	Animé	100.0%	Gigue A	92.3%
	Couplet 13	66.7%	Lentement	100.0%	Gigue B	100.0%
	Couplet 18	66.7%	Gravement	66.7%	Prelude	71.4%
	All	55.6%	All	88.9%	All	85.2%
2nd	Couplet 12	83.3%	Animé	85.7%	Gigue A	85.7%
	Couplet 13	83.3%	Lentement	85.7%	Gigue B	100.0%
	Couplet 18	83.3%	Gravement	71.4%	Prelude	66.7%
	All	83.3%	All	61.9%	All	73.7%
3rd	Couplet 12	83.3%	Animé	66.7%	Gigue A	83.3%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	83.3%
	Couplet 18	83.3%	Gravement	100.%0	Prelude	83.3%
	All	88.9%	All	77.8%	All	70.8%
4th	Couplet 12	100.0%				
	Couplet 13	83.3%				
	Couplet 18	66.7%				
	All	72.2%				
All	Couplet 12	70.8%	Animé	73.7%	Gigue A	81.3%
	Couplet 13	66.7%	Lentement	68.4%	Gigue B	84.2%
	Couplet 18	70.8%	Gravement	73.7%	Prelude	57.9%
	All	**69.4%**	**All**	**61.4%**	**All**	**67.1%**

reverberance in one room would change the nature of the decay of a staccato note. However, this was likely minimized to some degree by combining the audio frames into larger texture windows.

Restricting the training of the SVMs to only the tempo features still resulted in fairly high accuracy (see table 5.4). There is generally only a slight reduction compared to the feature set including tempo and intensity features but overall the results are fairly similar suggesting either that the dynamics between the two rooms are not very different or that the intensity

feature is correlated with the tempo. Overall, the results suggest that there are some measurable differences in tempo as a function of the room, however further analysis is needed to see precisely what these differences are and whether they are perceivable by a listener. Chapter 7 examines the question of the perceptibility of these changes more directly.

This also shows that the combined tempo and intensity features, and even the tempo features by themselves, are fairly discriminative features for describing musical performances. Furthermore, their direct musical meaning and easy explainability makes these features even more convenient for use in further analysis.

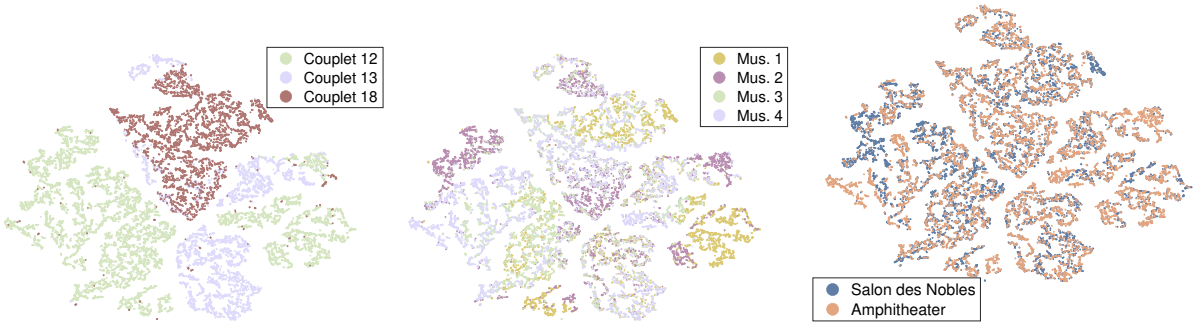
5.4 Individual musician results

The results for individual musicians will only be analyzed using intensity and tempo features, as these demonstrated high distinguishability and a certain level of robustness against the room sound in the recordings.

In order to determine the significance of the differences of the tempo and intensity between the two rooms, Friedman tests were performed with repetitions as blocks and the acoustics as group variable. A Friedman test was performed for each musician and composition individually on both note-level tempo and intensity (as measured by RMS). The p-values from the Friedman tests were calculated and are reported for each musician in the following sections; p-values with a significance level below .05 are shown in bold, while those with a significance level below .01 are denoted with an asterisk (*****), and those with a significance level below .001 are denoted with two asterisks (******).

In addition to examining the objective performance features, responses from the open-ended questions from the questionnaire which asked about specific adjustments made to their performance will be reported. These questions ask about the musician's impression of the room and how their playing may have been impacted by it and other factors. These responses can help to contextualize the objective measurements and assist in determining whether any measured differences in playing style between the two rooms is the result of an intentional strategy by the performer to adapt to the acoustics, or just the result of normal variation from one performance to another.

5.4.1 Violists



(a) (b) (c)

Figure 5.3 t-SNE projection of intensity and tempo features of viol performances separated by different classes: (a) composition, (b) performer, and (c) performance space.

To visualize the discriminability of the tempo and intensity features for the violists, a t-dstributed stochastic neighbor embedding (t-SNE) projection of the tempo and intensity features for all viol performances was made. This method was developed by Maaten and Hinton (2008) as an improved way to visualize high-dimensional data in two or three dimensions. Similarly to Stochastic Neighbor Embedding (Hinton and Roweis, 2002), t-SNE converts euclidean distances between data points in high dimensional space into conditional probabilities that represent similarities between data points. However, t-SNE aims to improve upon the tendency of Stochastic Neighbor Embedding to produce visualizations which tend to crowd toward the middle of the axes, resulting in visualizations that are more capable of representing structure of large data sets. This projection reveals the class separability of the compositions (fig. 5.3a), musicians (fig. 5.3b), and performance space (fig. 5.3c). As expected, these features allow clear discrimination between the different compositions. Additionally, they are quite capable of highlighting differences among the playing styles of the different musicians. The separation between the two performance spaces is evident in certain localities but much less distinct.

5.4.1.1 Violist 1

Violist 1's response to the open-ended questions (see table 5.5) indicate a strong preference for the Salon des Nobles over the amphitheater. Not only did they think that "the acoustics are absolutely perfect" but they also seemed impressed and influenced by the visuals and historical

Table 5.5 Responses of violist 1 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	This space is made for playing the viol. It was an immense pleasure to play in this place, the acoustics are absolutely perfect, the tone reproduced to the highest degree - an unforgettable experience.	I was a little bothered by the light. The color possibilities are quite huge and the dimensions of the room very pleasant. I was also a little tense, remembering a concert a long time ago in this same place on a Sunday morning at 11am when I was very nervous!
Was your performance influenced by the way you felt? If yes, in what way?	Being in this inspiring place made me appreciate every note. The sight of the magnificent adornment gave me plenty of new ideas while performing the pieces.	I had a certain nostalgia for the experience to end and I was perhaps not as inspired as at Versailles.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	I let myself be carried away by the moment.	Yes, as the resonance in the treble = almost zero, I tried to do the maximum without overdoing it I believe.

nature of the Salon des Nobles, stating that it gave them "plenty of new ideas while performing the pieces." By contrast, the violist seemed distracted by the colors within and size of the amphitheater. They also complained of a lack of resonance in the treble in the amphitheater.

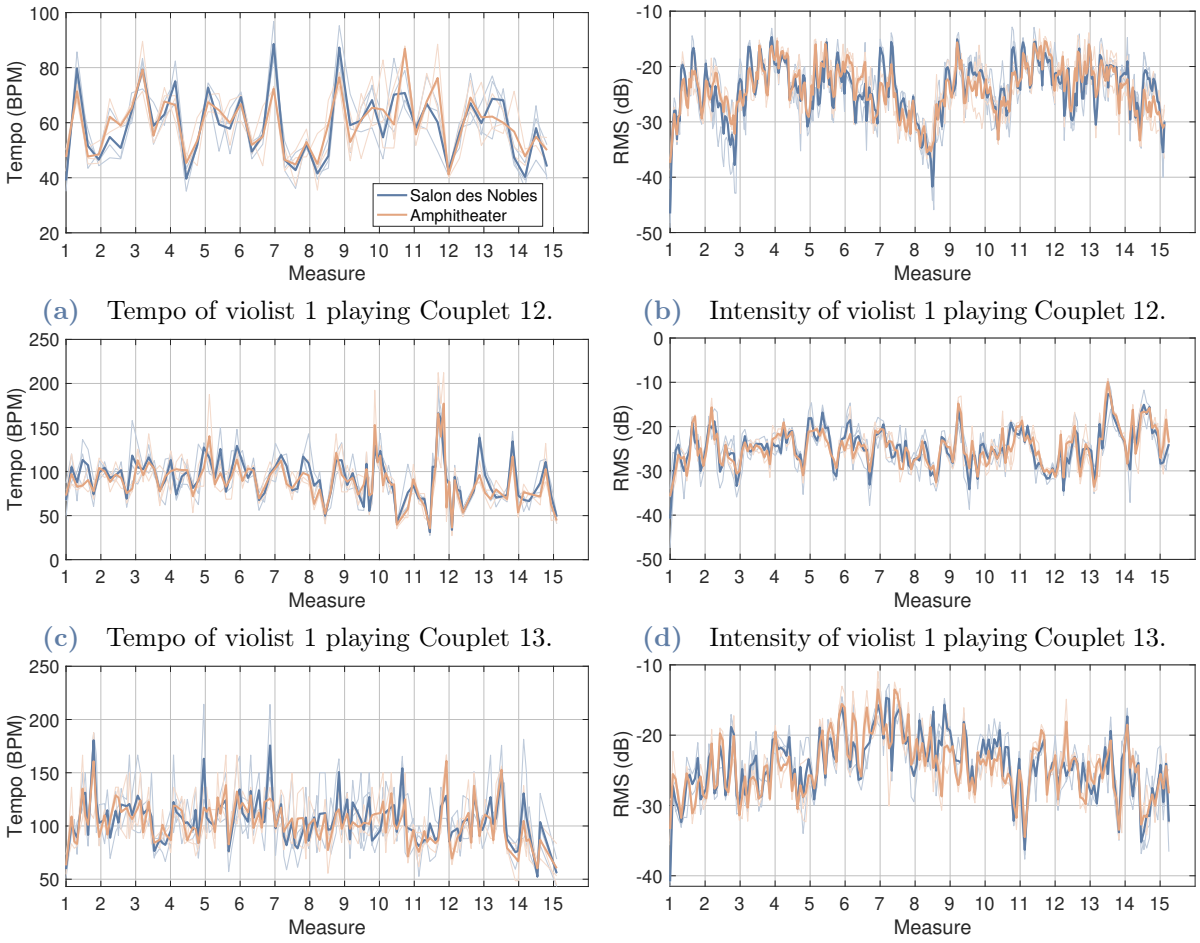
While this musician did not give many specifics as to how their playing style changed as a result of these strikingly different impressions of the two rooms, it is clear that they were more at ease in the Salon des Nobles, and therefore were likely better able to perform. This musician's response to the two rooms provides evidence that other external factors of a performance space, aside from the acoustics, can influence one's performance, although in this case the two main factors (the visuals and the acoustics) seem to be covariant.

Table 5.6 Friedman test p-values for tempo and intensity features of performances by violist 1.

Feature	Couplet 12	Couplet 13	Couplet 18
Tempo	.113	**.002***	.173
RMS	.403	.869	.167

The results of the Friedman tests for violist 1 (see table 5.6) showed a significant difference in tempo for one piece, Couplet 13 ($p = .002$). Time-series plots of the note-level tempo of the performances of this piece can be seen in fig. 5.4c. Some noticeable differences occur, such as in the last three measures where the tempo variations at the end of the measures are greater in the Salon des Nobles. This also appears to be true in moments of the other two compositions, Couplet 12 and Couplet 18 (see figs. 5.4a and 5.4e), but is perhaps less consistent across the repetitions.

No significant differences in RMS were found as a function of the room.



(a) Tempo of violist 1 playing Couplet 12.

(b) Intensity of violist 1 playing Couplet 12.

(c) Tempo of violist 1 playing Couplet 13.

(d) Intensity of violist 1 playing Couplet 13.

(e) Tempo of violist 1 playing Couplet 18.

(f) Intensity of violist 1 playing Couplet 18.

Figure 5.4 Tempo and RMS plots of all performances by violist 1. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.1.2 Violist 2

Table 5.7 Responses of violist 2 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	Very pleasing	This space had a very positive impact on my playing. The feeling was very good.
Was your performance influenced by the way you felt? If yes, in what way?	A little.	Probably a lot because [my emotional state] is very unstable at the moment.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	Not really.	I didn't need a conscious adjustment.

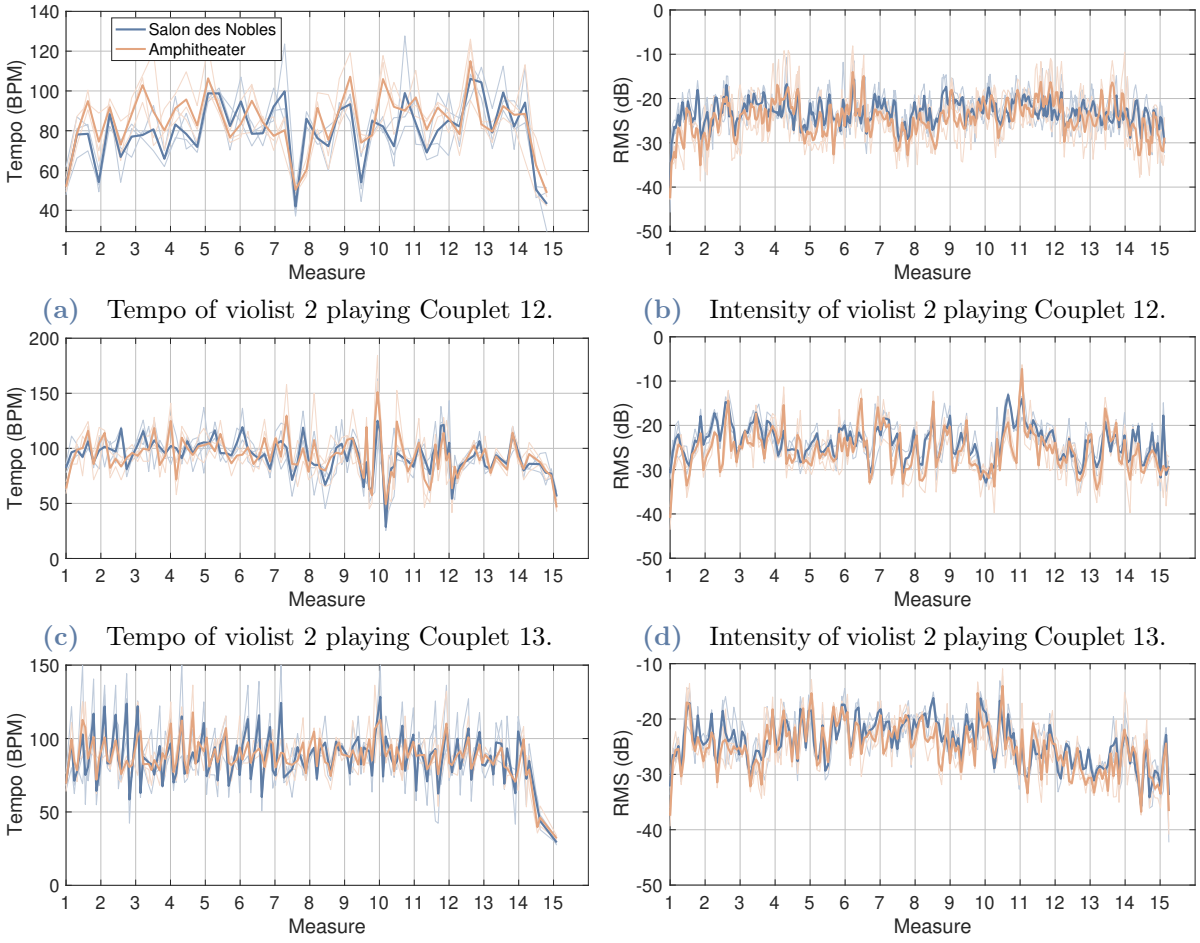
The responses to the open-ended questions by violist 2 do not give much insight into their playing or their impressions of the space. While they said the Salon des Nobles was "very pleasing", they also said the amphitheater had a "very positive impact on my playing." They mention no specific adjustments made to their playing.

Table 5.8 Friedman test p-values for tempo and intensity features of performances by violist 2.

Feature	Couplet 12	Couplet 13	Couplet 18
Tempo	$<$ **.001****	.554	.375
RMS	$<$ **.001****	$<$ **.001****	**.002***

One piece, Couplet 12, performed by violist 2 showed a highly significant difference in tempo ($p < .001$) between the two rooms. The time-series plot (see fig. 5.5a) shows notable differences between the two rooms although these differences are difficult to characterize.

All three pieces showed a significant difference among performances in the two rooms in terms of intensity with the first two pieces (Couplets 12 and 13) showing a highly significant difference ($p < .001$) with the third having a p-value of .002. Observing the time-series data in figs. 5.5b, 5.5d and 5.5f differences are noticeable but again difficult to characterize succinctly.



(a) Tempo of violist 2 playing Couplet 12.

(b) Intensity of violist 2 playing Couplet 12.

(c) Tempo of violist 2 playing Couplet 13.

(d) Intensity of violist 2 playing Couplet 13.

(e) Tempo of violist 2 playing Couplet 18.

(f) Intensity of violist 2 playing Couplet 18.

Figure 5.5 Tempo and intensity plots of all performances by violist 2. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.1.3 Violist 3

Violist 3, like violist 1, also felt positively about the acoustics in the Salon des Nobles, saying it was "the best acoustic" for the music of Marin Marais (the composer of the viol pieces). They also claimed that the acoustics of the Salon des Nobles allowed them to play with "more nuance" although it is not clear which performance dimension this nuance was applied to. By contrast, violist 3 stated that they had to adapt to the lack of resonance in the amphitheater which was similar to the response of violist 1.

Table 5.9 Responses of violist 3 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

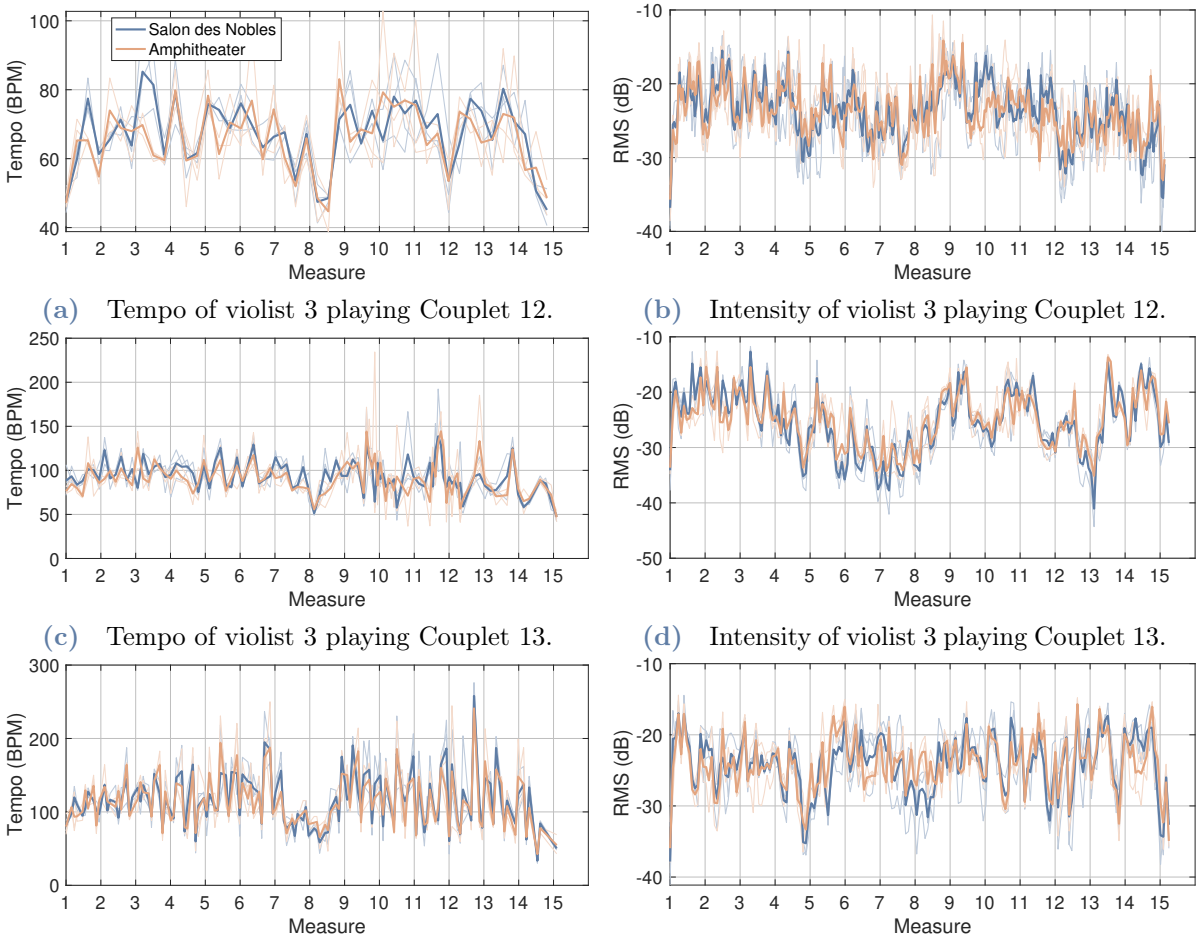
Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	Versailles "home to" Marin Marais is the best acoustic / framing for his music. What to say — inspiring.	Fairly neutral. Almost like a studio.
Was your performance influenced by the way you felt? If yes, in what way?	Very moved—always as a New Yorker to be able to play in the footsteps of Marin Marais.	I performed in this room when it opened — nostalgia.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	Yes — more nuance as the room responds.	Not much resonance so [I was] aware of this and therefore adapt[ed] to these considerations.

Table 5.10 Friedman test p-values for tempo and intensity features of performances by violist 3.

Feature	Couplet 12	Couplet 13	Couplet 18
Tempo	$<$ **.001****	$<$ **.001****	.221
RMS	.205	.843	.491

Highly significant differences in tempo were found in performances between the two rooms for the first two pieces, Couplet 12 and 13 ($p < .001$). Local differences can be found in the time-series tempo figures (see figs. 5.6a and 5.6c), but it is not clear if the musician was adhering to a specific strategy.

No significant differences were found among the performances in the two rooms in terms of the intensity.



(a) Tempo of violist 3 playing Couplet 12.

(b) Intensity of violist 3 playing Couplet 12.

(c) Tempo of violist 3 playing Couplet 13.

(d) Intensity of violist 3 playing Couplet 13.

(e) Tempo of violist 3 playing Couplet 18.

(f) Intensity of violist 3 playing Couplet 18.

Figure 5.6 Tempo and intensity plots of all performances by violist 3. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.1.4 Violist 4

Violist 4 had a notable preference for the Salon des Nobles according to their responses to the open-ended questions (see table 5.11). This environment, they stated, was "literally adapted" to their instrument. Furthermore, they said that the reverberation in the Salon des Nobles allowed them to play with "more straightforward or lighter articulations" which align with the expectations of historical baroque performance practice (see section 2.1.3).

By contrast violist 4 explained that the size of the amphitheater influenced the level of

Table 5.11 Responses of violist 4 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	The size of the room allows you to feel more free, perhaps less intimidating. The instrument being baroque, I feel better in an environment "literally adapted" to my instrument.	The size of the venue influenced the level of dynamics I used, and the reverb influences the tempi. I always need some adaptation time = the first few minutes of adjustment are difficult then ultimately I'm very comfortable.
Was your performance influenced by the way you felt? If yes, in what way?	The stress of the new place always affects the first takes, I still need a little time to adapt. In general, I gradually build up the tempos but always start a little stressed.	My emotional state still influences the way I play, the stress of playing in front of people again and recording made the first few takes more acoustically complex = we've lost the habit of hearing each other.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	Once the reverb was heard / taken into account I was able to play more slowly in general with more straightforward or lighter articulations depending on the movements. The more I hear myself, the more I try to reign in certain details.	I adjusted to the reverberation, adapted the tempi and the length of the notes as well as the dynamics, all this according to the response of the room.

dynamics used and that its reverberation influenced the tempi. However, they do not elaborate precisely on how these parameters were adjusted in response. Another interesting statement made by violist 4 is that they adjusted the length of their notes in response to the reverberation. While they did not specify whether they shorted or lengthened the notes, one strategy which has been previously noted is that musicians attempt to lengthen notes in environments with little reverberation as a way to compensate for the lack of acoustical decay (see section 2.3). Furthermore, this practice of lengthening notes is contrary to a historical baroque performing style which values a detached or separated articulation over a legato or connected style (see section 2.1.3 for more on historically informed baroque performance practice). In other words, this violist may have been compelled to play in a less baroque appropriate way in the amphitheater due to its acoustics.

Violist 4 showed highly significant differences in tempo between the two rooms for all three pieces ($p < .001$). The time-series tempo data for Couplet 12 (see fig. 5.7a) shows that, in general the tempo in the amphitheater was faster in the first half of the piece then slowed down

Table 5.12 Friedman test p-values for tempo and intensity features of performances by violist 4.

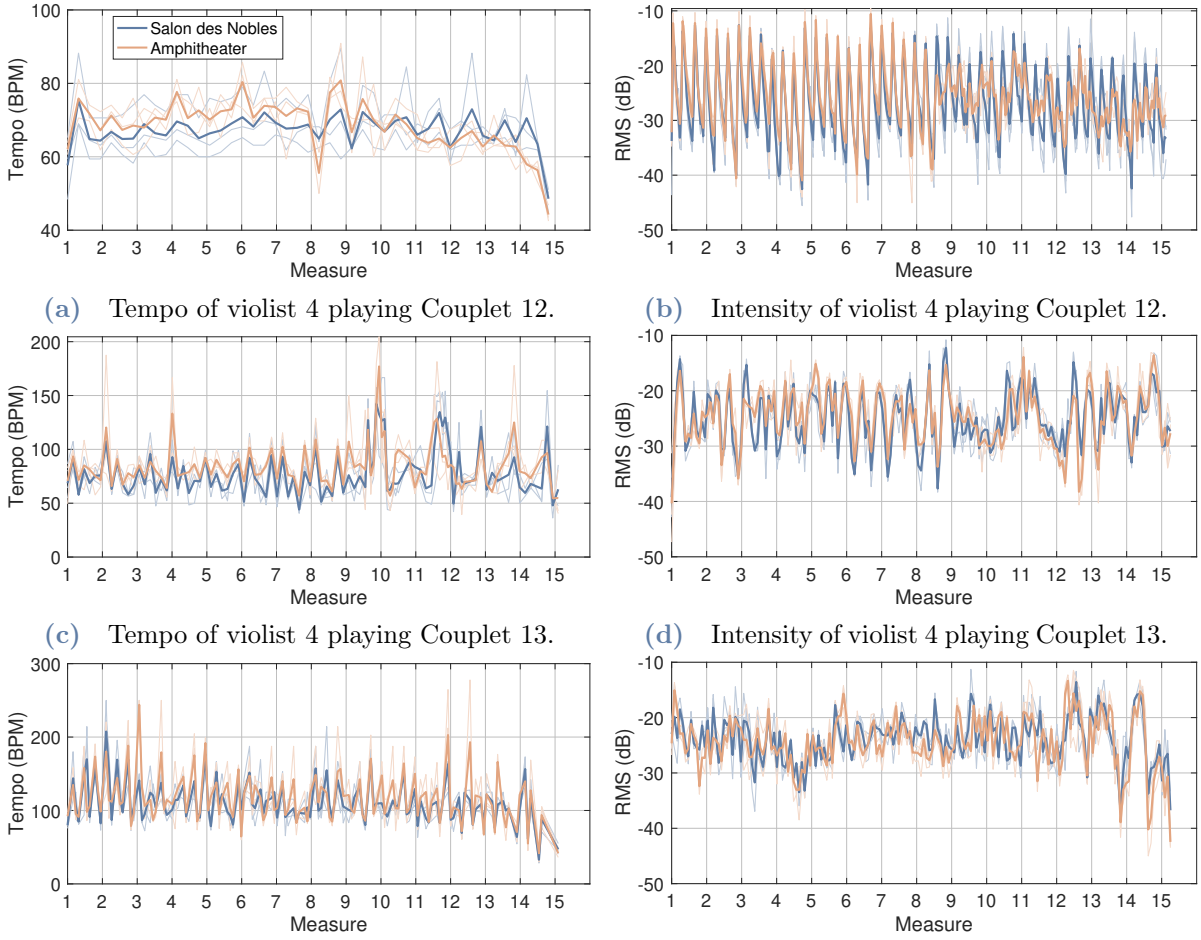
Feature	Couplet 12	Couplet 13	Couplet 18
Tempo	$<$ **.001****	$<$ **.001****	$<$ **.001****
RMS	$<$ **.001****	.218	.144

in the second half of the piece. This is due to the fact that, in the amphitheater, this musician played chords in one bow stroke for the first half of the piece, then switched to arpeggiating them in the second half of the piece which resulted in a slower tempo. In the Salon des Nobles, however, they played the chords in one bow stroke for the entirety of the piece. Therefore, the tempo is much more steady for the entire performance in the Salon des Nobles, while it varies in the second half of the performances in the amphitheater.

This musician did not mention such a contrast in playing in their comments so while this change in arpeggiation is largely responsible for the measured significant differences (see table 5.12), it does not appear to be an intentional adjustment in response to the room's acoustics.

In general, in Couplet 13, it appears that the tempo in the Salon des Nobles was slightly slower, especially in the middle of the piece (see fig. 5.7c). The tempo differences in Couplet 18, while still highly significant, can not be easily summarized based on the time-series data (see fig. 5.7e).

In terms of intensity, performances of Couplet 12 were different at a highly significant level ($p < .001$). The time-series intensity data for this piece (see fig. 5.7b) is highly distinct when compared to the other musicians. This is due to the playing style described above (playing chords in one bow stroke rather than arpeggiating). It is easy to see when the musician switched to arpeggiating in the amphitheater performances in the second half of the piece, where there is a noticeable difference in the intensity among performances in the two rooms.



(a) Tempo of violist 4 playing Couplet 12.

(b) Intensity of violist 4 playing Couplet 12.

(c) Tempo of violist 4 playing Couplet 13.

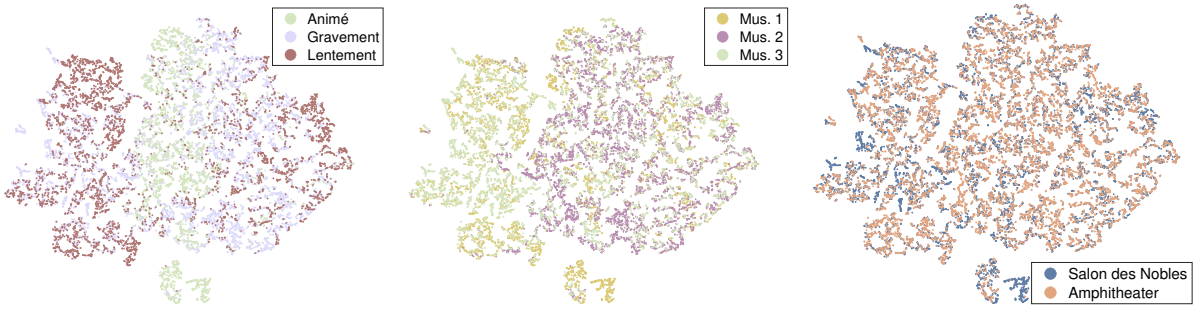
(d) Intensity of violist 4 playing Couplet 13.

(e) Tempo of violist 4 playing Couplet 18.

(f) Intensity of violist 4 playing Couplet 18.

Figure 5.7 Tempo and intensity plots of all performances by violist 4. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.2 Flutists



(a) (b) (c)

Figure 5.8 t-SNE projection of intensity and tempo features of flute performances separated by different classes: (a) composition, (b) performer, and (c) performance space.

While the separation is evident for flute pieces, as shown by the t-SNE projection in fig. 5.8a, it is not quite as clear as for the viol pieces. The projection indicating the different musicians, in fig. 5.8b, likewise shows some moderate separation. As was the case for the violists, the projection showing the different rooms (fig. 5.8c) is quite mixed, however, there are portions of the data that show quite clear separation.

5.4.2.1 Flutist 1

Flutist 1 felt positively about the acoustics of the Salon des Nobles, stating that in the Salon des Nobles they had the "impression of a full space." By contrast, they said their playing "seemed to be very exposed" in the amphitheater. Consequently, the flutist said they reveled in the ends of phrases in the Salon des Nobles while in the amphitheater they extended the duration of notes to fill the entire duration of the written note value. This is similar to strategies previously described for adapting to rooms with insufficient reverberance. Furthermore, since historical baroque performance practice emphasizes a detached playing style, this indicates that this flutist was adopting a less historically appropriate baroque style in the amphitheater as a consequence of its acoustics (see section 2.1.3 for more on historical baroque playing style).

Highly significant differences in tempo were found in performances of all pieces between the rooms ($p < .001$) for flutist 1. Observing the time series data in figs. 5.9a and 5.9c, it appears that the flutist played slower in the Salon des Nobles for these two pieces, supporting some earlier findings that musicians tend to player slower in more reverberant spaces. This observation

Table 5.13 Responses of flutist 1 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	This space made me want to imprint my sound more as if to be able to extend it to every corner of the room. I like to be able to appreciate the ends of phrases, the ends of sounds. [I had the] impression of a full space.	My playing seemed to be very exposed where every detail could be heard. I also felt like I needed to project further, to hold on until the very end of the note values in order to reach the back of the room.
Was your performance influenced by the way you felt? If yes, in what way?	Emotional state a little stressed by unknown conditions this affects breathing taken too high and too short.	I was present on stage and tried to appreciate the characteristics of the hall, to pay attention to it. The first pass, I always stay more focused on myself, on my notes, then as time goes by, you loosen up and the attention to your surroundings increases.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	Adjustment to the level of vertical space. More anchoring to the ground and more projection upwards. More effort to put my stamp on the sound for a greater immersion in the space.	I think I accentuated the character of the pieces, sharper for Animé maybe slower of Lentement. I played very in front of me with horizontal projection & note values held to the very end.

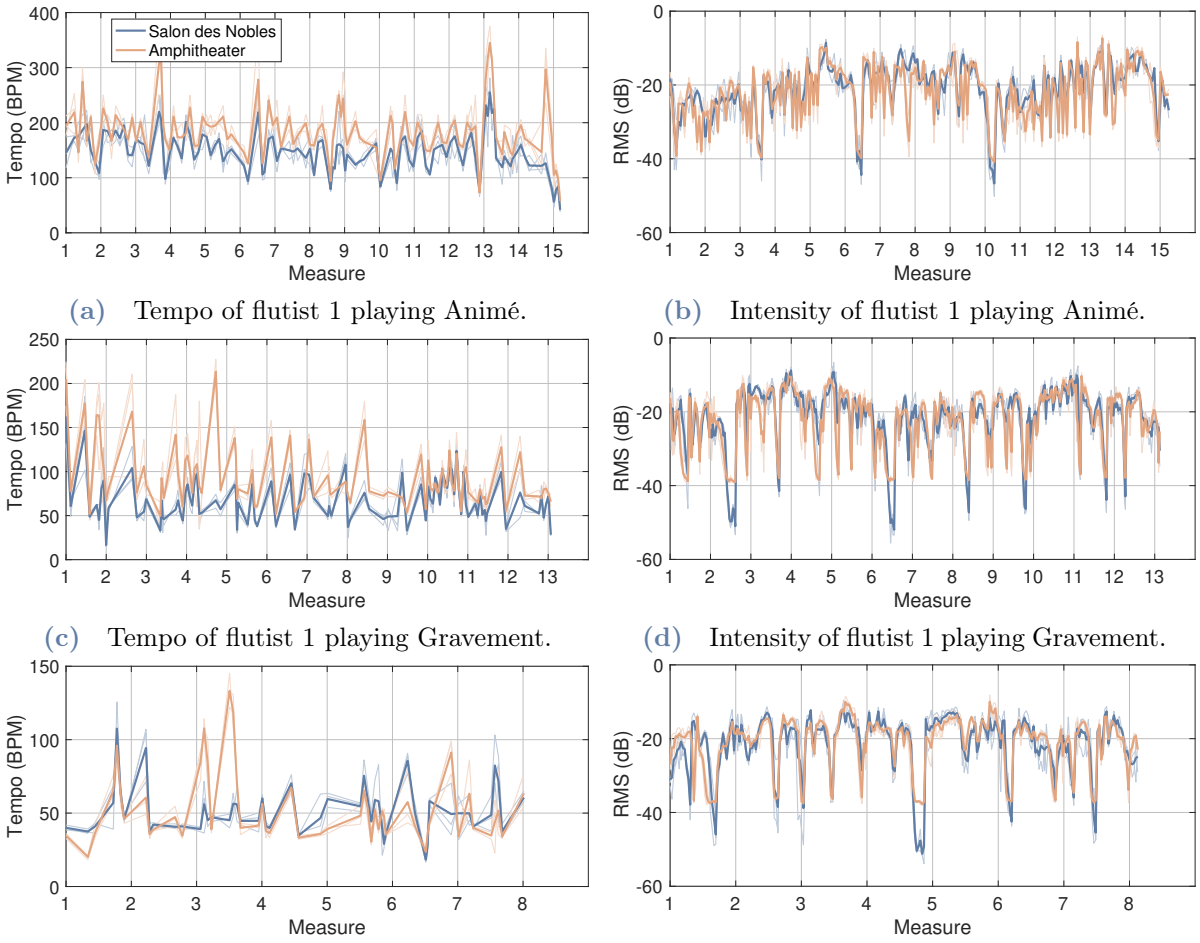
Table 5.14 Friedman test p-values for tempo and intensity features of performances by flutist 1.

Feature	Animé	Gravement	Lentement
Tempo	$<$ **.001****	$<$ **.001****	$<$ **.001****
RMS	**.025**	.262	.643

also aligns with the claims made by flutist 1 about how they adapted their playing in response to the acoustics. The attempt to hold the notes to their full duration in the amphitheater may have the consequence of speeding up the tempo as the psychological effect of wanting to minimize silence between notes may cause one to articulate the next note slightly earlier than otherwise, resulting in a higher overall tempo. The tempo differences in Lentement are also visible in the time-series data (fig. 5.9e), although precise differences can not be succinctly described.

This flutist demonstrated significant differences in terms intensity for the first piece, Animé

($p = .025$) though the time series data do not easily reveal what these differences are (see fig. 5.9b). According to the flutist's comments in table 5.13, they attempted to extend the duration of each note to its full written value in the amphitheater. This may be the reason the local minima in the amphitheater are not as low as those in the Salon des Nobles. These troughs are presumably moments the flutist briefly paused to take a breath or assert a phrase boundary but the lengthening of the notes in the amphitheater likely abbreviated this boundary resulting in a less acute decay.



(a) Tempo of flutist 1 playing Animé.

(b) Intensity of flutist 1 playing Animé.

(c) Tempo of flutist 1 playing Gravement.

(d) Intensity of flutist 1 playing Gravement.

(e) Tempo of flutist 1 playing Lentement.

(f) Intensity of flutist 1 playing Lentement.

Figure 5.9 Tempo and intensity plots of all performances by flutist 1. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.2.2 Flutist 2

Table 5.15 Responses of flutist 2 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

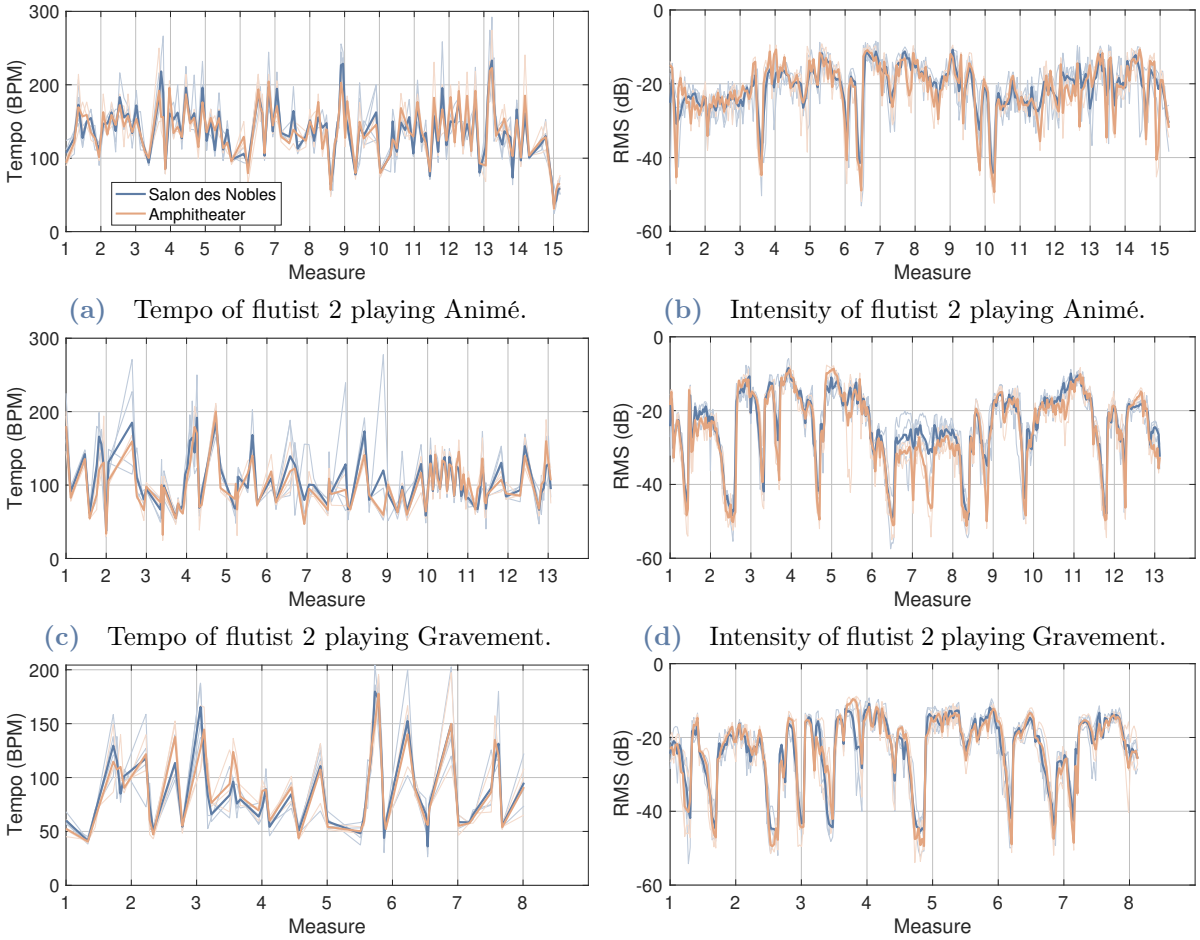
Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	Pleasant room. Suitable for the music.	Neutral.
Was your performance influenced by the way you felt? If yes, in what way?	Neutral.	Nothing to report.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	No, but pleasure to hear the response of the room.	Played longer because of the dryness of the room.

Similar to many of the other musicians, flutist 2 seemed to appreciate the acoustics of the Salon des Nobles more than the amphitheater. They asserted that the Salon des Nobles was "pleasant" and "suitable for the music" whereas they felt "neutral" about the amphitheater. In adapting to the different acoustics, they mentioned a strategy similar to the previous flutist, stating that they "played longer because of the dryness of the room." This, again is contrary to the accepted historical baroque playing style, indicating that the playing style in the Salon des Nobles was more baroque-appropriate than in the amphitheater due to the difference in acoustics in the two rooms.

Table 5.16 Friedman test p-values for tempo and intensity features of performances by flutist 2.

Feature	Animé	Gravement	Lentement
Tempo	.633	$<$ **.001****	.287
RMS	.677	$<$ **.001****	.977

Flutist 2 demonstrated highly significant differences, according to the Friedman tests, in their performances of Gravement between the two rooms both in terms of tempo and intensity ($p < .001$). Their self-reported adaptation strategy, however, is not as evident in the observed data as was the case for flutist 1.



(a) Tempo of flutist 2 playing Animé.

(b) Intensity of flutist 2 playing Animé.

(c) Tempo of flutist 2 playing Gravement.

(d) Intensity of flutist 2 playing Gravement.

(e) Tempo of flutist 2 playing Lentement.

(f) Intensity of flutist 2 playing Lentement.

Figure 5.10 Tempo and intensity plots of all performances by flutist 2. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.2.3 Flutist 3

The questionnaire responses for flutist 3 (see table 5.17) indicate that they felt fairly positively about the acoustics in both rooms. In the Salon des Nobles they said the acoustics seemed "simple" and that the room response was "very pleasant." In the amphitheater, they said the space sounded "quite generous." Similar to the other flutists, flutist 3 mentions that they attempted to "make the phrases longer" in the amphitheater. This provides strong evidence that the flutists feel as if they must change their natural playing style in order to compensate for

Table 5.17 Responses of flutist 3 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

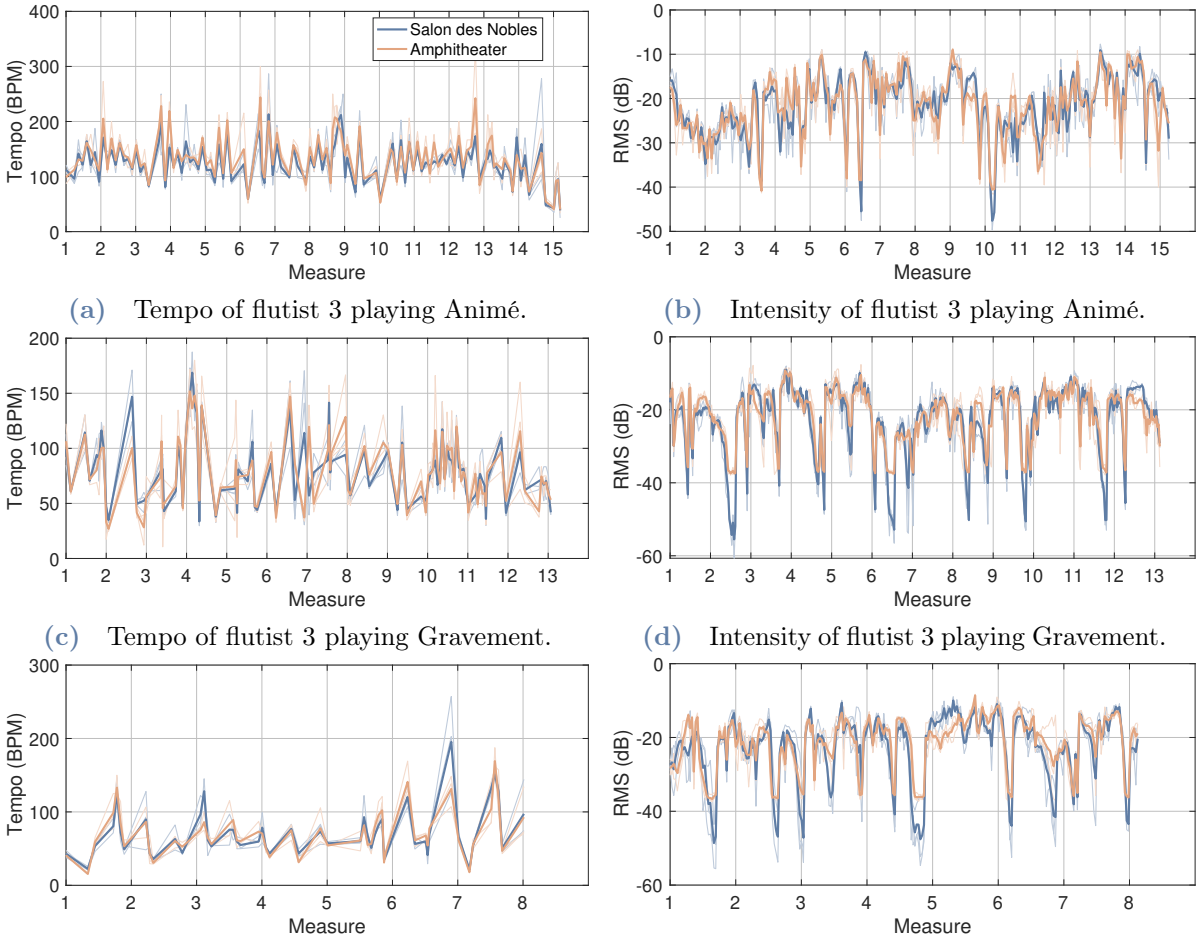
Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	The acoustics seem very simple. The sound is alive, the response is very pleasant. The fact of playing "in situ" in the salon of the château is very pleasant, stimulating for the imagination.	The space is large, sounds quite generous, it's a fairly "neutral" room (in decor) so it's quite easy to concentrate. There is a certain ease in finding yourself in a fairly traditional concert venue.
Was your performance influenced by the way you felt? If yes, in what way?	There is an emotion attached to being in the château almost alone. This influenced my playing in the sense of seeking an elegance that responds to the place.	I felt pretty relaxed, especially landing the second and third takes, so the playing was rather easy.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	I think I slowed down the tempo a bit to allow time to hear the sound resonate and bring the long notes to life.	Feeling like I was in a big space led me to dig deeper into the dynamics for the 2nd and 3rd takes, to support the sound more, to make the phrases longer.

acoustical differences in the amphitheater. Furthermore, this change in playing style results in a less baroque-appropriate style in the modern hall compared to the baroque hall, according to the musicological consensus on historically informed baroque performance practice (see section 2.1.3).

Table 5.18 Friedman test p-values for tempo and intensity features of performances by flutist 3.

Feature	Animé	Gravement	Lentement
Tempo	$<$ **.001****	.711	.475
RMS	.914	.514	.108

There was a highly significant difference found between the performances of flutist 3 in the two rooms of the piece Animé in terms of tempo ($p < .001$). Despite claiming that they "slowed down the tempo a bit" in the Salon des Nobles, the tempo plot for this piece (fig. 5.11a) does not clearly show an overall slower tempo in the Salon des Nobles.



(a) Tempo of flutist 3 playing Animé.

(b) Intensity of flutist 3 playing Animé.

(c) Tempo of flutist 3 playing Gravement.

(d) Intensity of flutist 3 playing Gravement.

(e) Tempo of flutist 3 playing Lentement.

(f) Intensity of flutist 3 playing Lentement.

Figure 5.11 Tempo and intensity plots of all performances by flutist 3. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.3 Theorbists

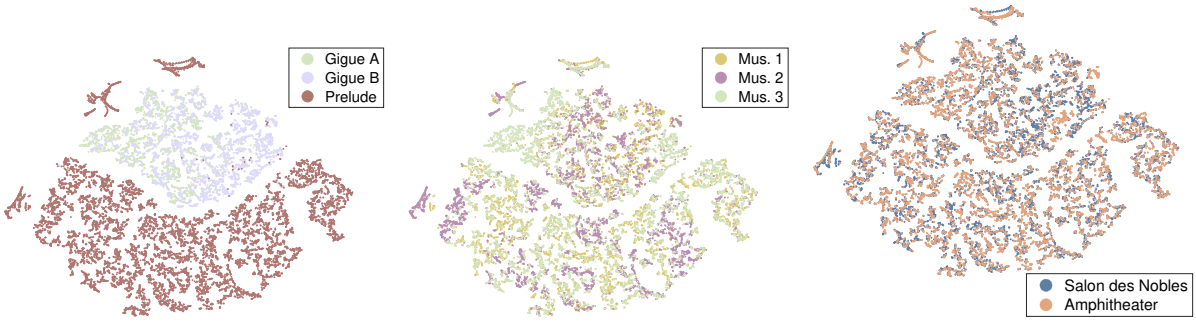


Figure 5.12 t-SNE projection of intensity and tempo features of the-orbo performances separated by different classes: (a) composition, (b) performer, and (c) performance space.

The t-SNE projection showing the separate compositions performed by the theorbists show clear separation of the different pieces. It should be noted that Gigue A and Gigue B are two sections of the same piece that were treated as two pieces for ease of analysis (see section 4.1.2.3) so the fact that there is not clear separation between these two pieces should not be considered an issue. The figure highlighting the musicians (fig. 5.12b) shows fairly moderate separation. And lastly, the projection showing the different rooms shows only localized portions of discriminability (fig. 5.12c).

5.4.3.1 Theorbist 1

According to their responses to the open-ended questions, Theorbist 1 clearly preferred the Salon des nobles (see table 5.19). They described the Salon des Nobles as a "nice room" while commenting that the amphitheater was an "unflattering room."

Some statements made by theorbist 1 align with some previous comments. They complained about the lack of decay in the amphitheater, saying "sounds don't last" there. As a result, they stated that the tempo was affected due to the "tendency to press on in order to fill the void", however, they also claimed that they intentionally fought against this tendency.

Highly significant differences in tempo were found for all pieces between the two rooms ($p < .001$). The overall tempo for Gigue A seems to be faster in the Salon des Nobles than in the amphitheater (see fig. 5.13a), running contrary to the accepted wisdom that musicians

Table 5.19 Responses of theorbist 1 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	Nice room, I'm fine.	Unflattering room — sounds don't last hence the fear of emptiness.
Was your performance influenced by the way you felt? If yes, in what way?	No impact.	In the middle of a pandemic, having not played in front of people for months = stage fright, strange experience.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	Nothing conscious, sensible to play the instrument.	With the short duration of the plucked strings, in a room with little reverberation, it is the timing that is affected = tendency to press on in order to fill the void. So I try not to do it all while doing it = complicated.

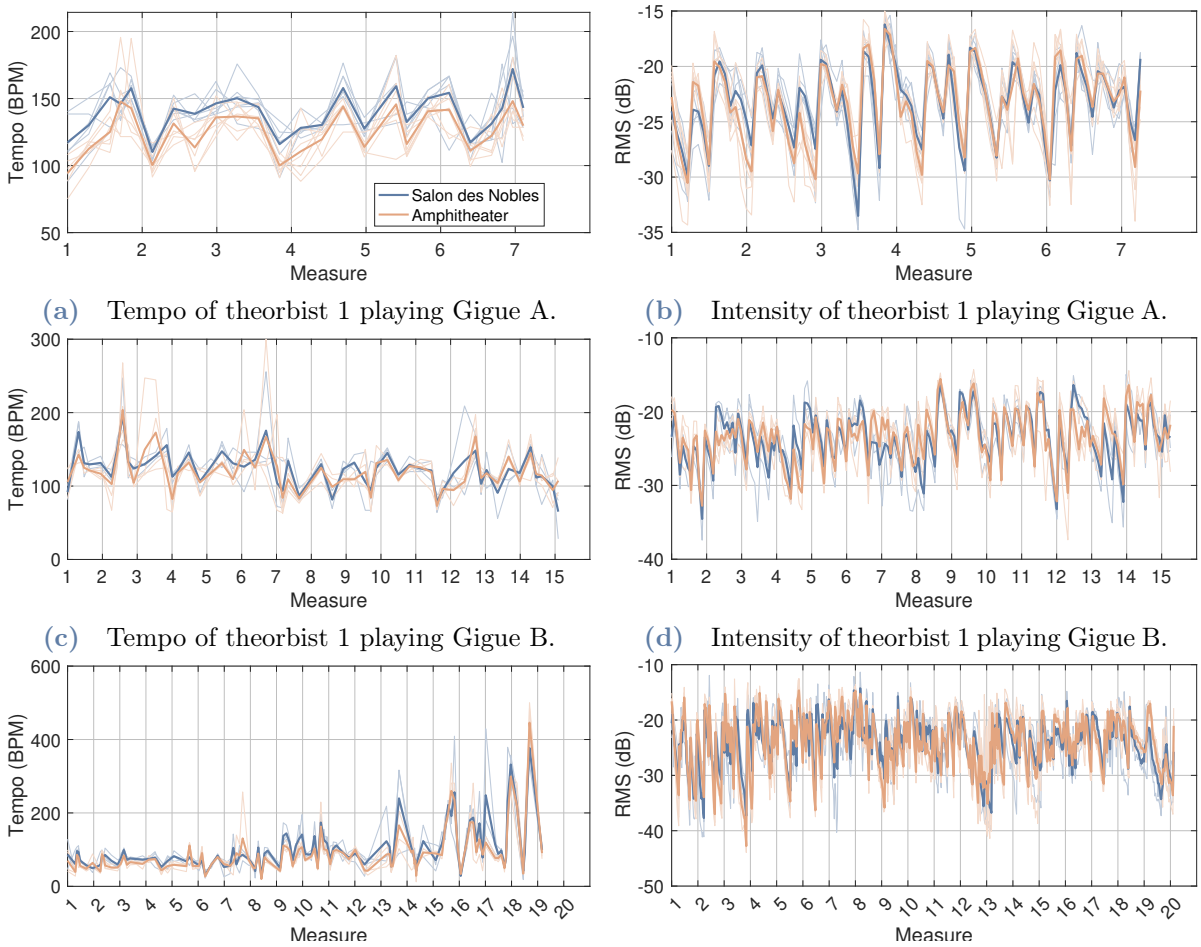
Table 5.20 Friedman test p-values for tempo and intensity features of performances by theorbist 1.

Feature	Gigue A	Gigue B	Prelude
Tempo	$<$ **.001****	$<$ **.001****	$<$ **.001****
RMS	.727	.440	.870

tend to play slower in more reverberant spaces (see section 2.3). This trend continues to some degree in both Gigue B and Prelude.

This also runs contrary to claims that they felt there was a tendency to speed up in the amphitheater due to the lack of acoustical support. One observation which may partially explain this is that this musician appeared to have been influenced by another musician (theorbist 3) that they first encountered in the Salon des Nobles. It is possible that theorbist 1 heard the much faster tempo used by theorbist 3 during the session of the Salon des Nobles and knowingly or unknowingly attempted to imitate this tempo, resulting in a slightly faster set of performances in the Salon des Nobles.

There were no significant differences found in the intensity of performances by theorbist 1 between the two rooms.



(a) Tempo of theorbist 1 playing Gigue A.

(b) Intensity of theorbist 1 playing Gigue A.

(c) Tempo of theorbist 1 playing Gigue B.

(d) Intensity of theorbist 1 playing Gigue B.

(e) Tempo of theorbist 1 playing Prelude.

(f) Intensity of theorbist 1 playing Prelude.

Figure 5.13 Tempo and intensity plots of all performances by theorbist 1. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.3.2 Theorbist 2

Theorbist 2 found the Salon des Nobles "very suitable" for the theorbo though would have preferred "a little more reverberation" according to their responses to the open-ended questions (see table 5.21). By contrast they felt that the amphitheater "did not have a lot of resonance" which led to them feeling "exposed" but that this also allowed them to hear everything clearly.

Like many other musicians, theorbist 2 described an effort to overcome the lack of resonance in the amphitheater whereas they were able to play as usual in the Salon des Nobles. Another

Table 5.21 Responses of theorbist 2 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

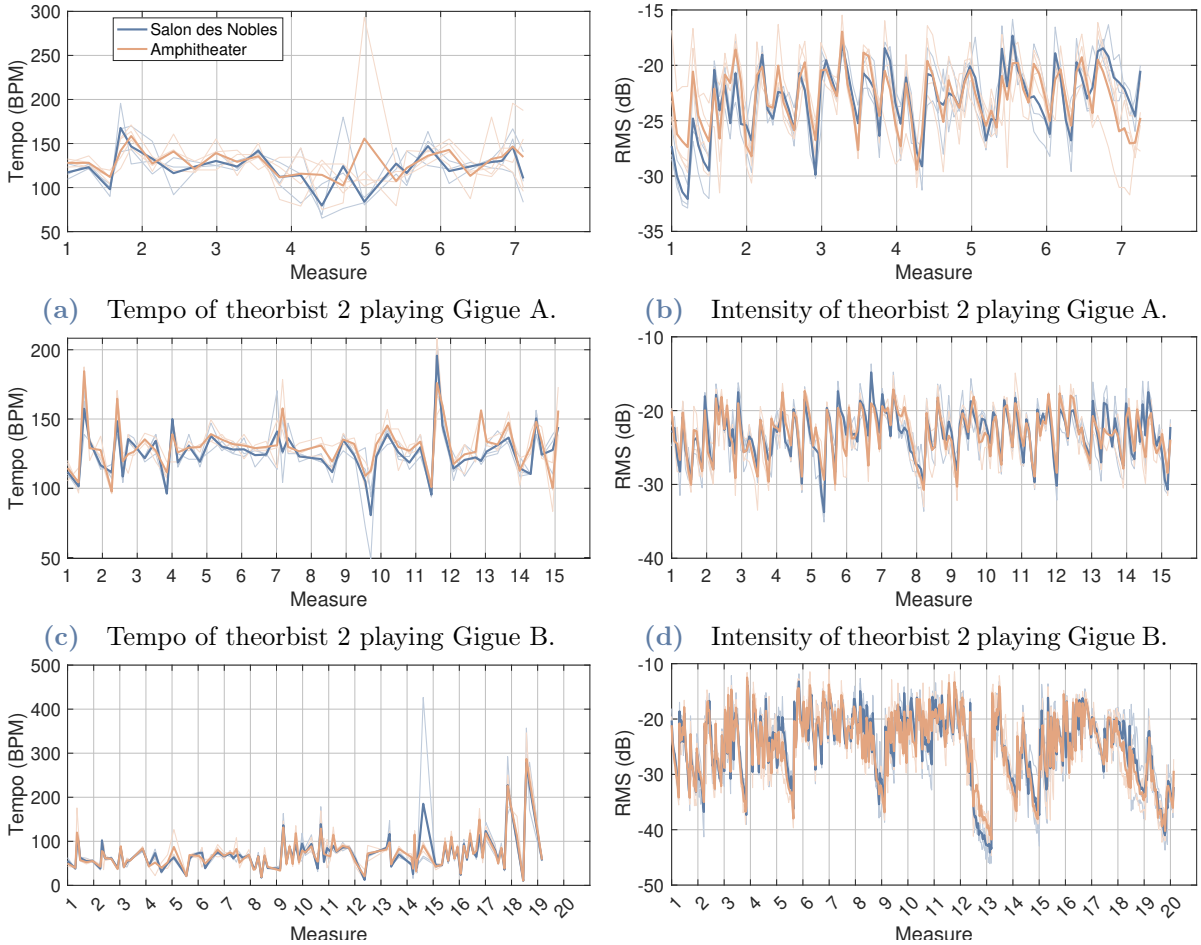
Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	I found the hall very suitable for the instrument the size of the hall was perfect, I would have liked a little more reverberation, but I liked it.	I felt like it did not have a lot of resonance and had a bit of a dry tone that wasn't excessive but I felt exposed. However, I heard everything clearly.
Was your performance influenced by the way you felt? If yes, in what way?	None.	It affected me moderately, I didn't sleep very well, and had little time to prepare repertoire.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	No, I played like I usually do.	Yes, I tried to make the instrument resonate more by using my fingers more towards the instrument top. I put more pressure on the strings to make the instrument vibrate more.

notable comment they made was they felt they had "little time to prepare [the] repertoire." This supports previous reports that the theorbists generally struggled to play their repertoire without mistakes.

Table 5.22 Friedman test p-values for tempo and intensity features of performances by theorbist 2.

Feature	Gigue A	Gigue B	Prelude
Tempo	**.002***	$<$ **.001****	.069
RMS	.970	.912	.061

Similarly to theorbist 1, theorbist 2 showed some significant differences in tempo, particularly for the first two pieces, Gigue A ($p = .002$) and Gigue B ($p < .001$). The differences are not very consistent or clear, however. There does appear to be a fairly large variety in tempo again, including some extreme values, probably due to the previously discussed difficulty level of the pieces chosen for the theorbo. No significant differences in intensity were found for any piece performed by theorbist 2.



(a) Tempo of theorbist 2 playing Gigue A.

(b) Intensity of theorbist 2 playing Gigue A.

(c) Tempo of theorbist 2 playing Gigue B.

(d) Intensity of theorbist 2 playing Gigue B.

(e) Tempo of theorbist 2 playing Prelude.

(f) Intensity of theorbist 2 playing Prelude.

Figure 5.14 Tempo and intensity plots of all performances by theorbist 2. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.4.3.3 Theorbist 3

Theorbist 3 seemed to enjoy the acoustics of both spaces according to the responses reported in table 5.23. However, even their positive claims about the acoustics of the amphitheater indicate their expectations were low, saying they were "pleasantly surprised" and that they were "very surprised to see that the room responds extremely well to the instrument." Their favorable opinion of the Salon des Nobles suggests it is largely due to the historical nature of the room, not merely the acoustics. Contrary to many of the other musicians, theorbist 3 felt there was

Table 5.23 Responses of theorbist 3 to the open-ended questions at the end of the acoustic rating questionnaire (see table A.1).

Question	Salon des Nobles	Amphitheater
How did this space impact your playing? How did it make you feel?	The pleasure of being in a salon at the Château de Versailles made me want to project these pieces, in order to pay homage to them in a way.	Pleasantly surprised because very good projection no need to force the instrument, which is an asset in a stressful situation.
Was your performance influenced by the way you felt? If yes, in what way?	Personal state of fatigue + as well as the fact of not having been able to play the instrument during the day has a strong impact on concentration and the impression of a comfortable handling of the instrument.	Playing "cold" increases the level of stress and therefore less comfort than with time to warm up beforehand. Moreover, this room being a real concert hall adds more "psychological pressure" than in the salon at Versailles for example.
Were you conscious of any adjustments to your performance due to the space? To what would you attribute this?	Yes and no, but more or less instinctively I try to rid as many small harmonics as possible compared to what the room offers as an acoustic space conducive or not to the development of harmonics, which in this specific case was very resonant.	Very surprised to see that the room responds extremely well to the instrument, therefore "positive" adjustment of muscle relaxation because [there is] no need to "force" on the acoustic sound which invites you to "let yourself be carried away."

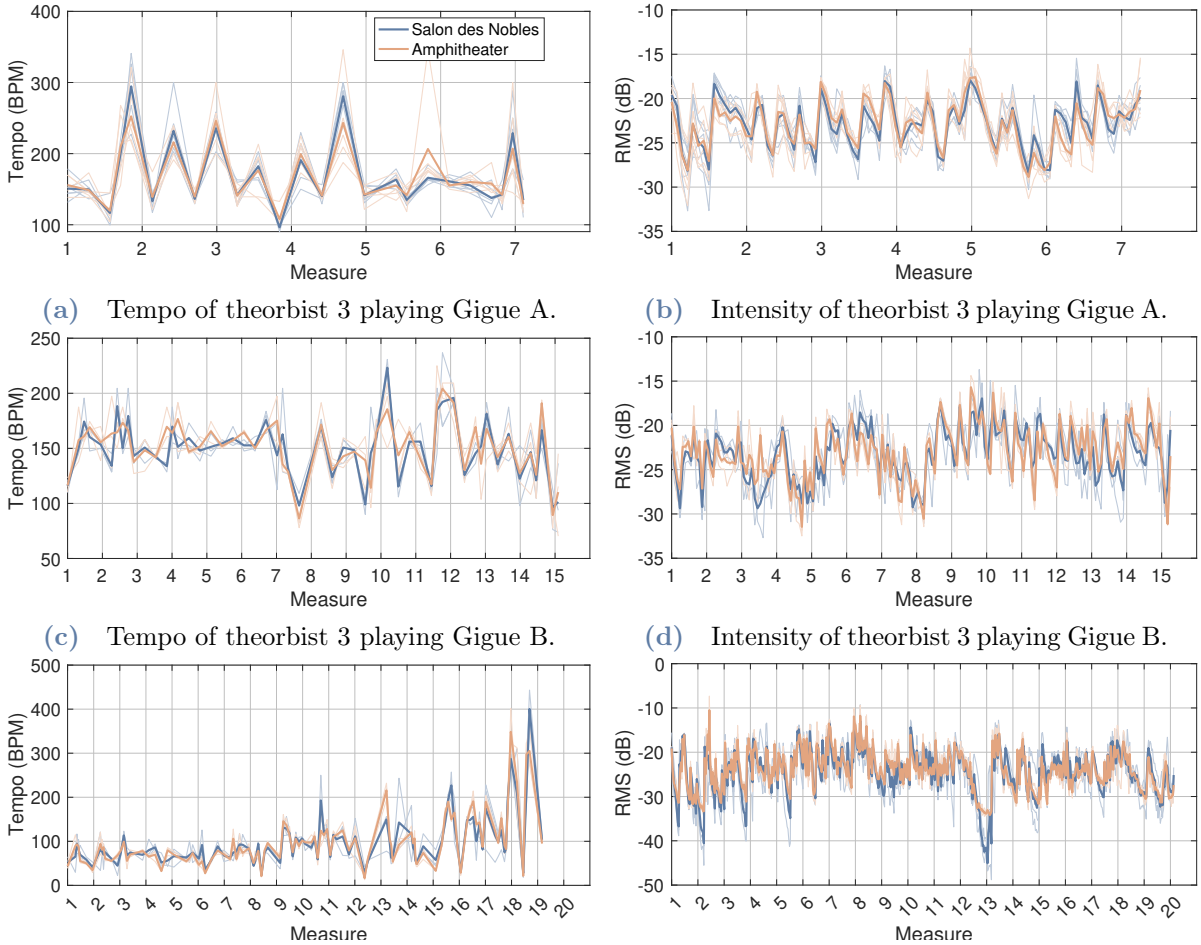
no need to force a sound in the amphitheater.

Table 5.24 Friedman test p-values for tempo and intensity features of performances by theorbist 3.

Feature	Gigue A	Gigue B	Prelude
Tempo	.482	**.011**	.899
RMS	.716	.445	.872

There was a significant difference in the performances of Gigue B in terms of tempo between the two rooms ($p = .011$). The differences, seen in fig. 5.15c, are not very consistent and difficult to interpret. The overall tempo variation seems to be less for this musician which correlates with their ability to perform these pieces more consistently across repetitions.

No significant differences in intensity were found among theorbist 3's performances between the two rooms.



(a) Tempo of theorbist 3 playing Gigue A.

(b) Intensity of theorbist 3 playing Gigue A.

(c) Tempo of theorbist 3 playing Gigue B.

(d) Intensity of theorbist 3 playing Gigue B.

(e) Tempo of theorbist 3 playing Prelude.

(f) Intensity of theorbist 3 playing Prelude.

Figure 5.15 Tempo and intensity plots of all performances by theorbist 3. The thick lines represent the means of all performances while the thin, semi-transparent lines represent the individual performances.

5.5 Discussion

In order to better understand how distinct the performance style was in one room compared to the other, SVM classifiers were trained on the extracted performance features and tasked with predicting which room the performances occurred in. The resulting accuracy of these classifiers was very high, suggesting a marked difference between the performances in the two different rooms (see table 5.2). However, much of this measured difference may not have been due to changes in playing style, but rather due to the sound of the room incidentally present in the

recorded signal. The presence of this sound likely had a significant influence on the timbre features and therefore strongly influenced the results of the classifiers.

In order to have a better idea of how the performances actually varied between the rooms, additional SVM classifiers were trained using two smaller feature subsets. One subset included the tempo and intensity features and one subset included only tempo features. The influence of the room sound may still be a factor for the feature set including intensity, but is likely less than the full feature set with all timbre features. Furthermore, these features have a fairly direct musical meaning making it much easier to interpret measured differences compared to harder-to-interpret timbre features. The resulting classification accuracies show a more realistic picture of the variety in playing style between the two rooms (see table 5.3). The accuracies of the data subsets that include all musicians and all pieces (within a single instrument class) predicted the room correctly around 2/3 of the time. The results of these same data subsets using only tempo features show similar accuracies ranging from about 61% to 69%. This suggests a more subtle range of playing styles between the two rooms. It also indicates that there were some individual adaptation strategies which were not shared by the entire group, a conclusion which has been found in several previous studies, as discussed in section 2.3 (Luizard, Brauer, et al., 2019; Schärer Kalkandjiev and Weinzierl, 2015; Amengual Gari et al., 2019).

Analyses of individual musicians was also undertaken by performing Friedman tests for each composition for both tempo and intensity features to see if any significant differences were measured between the two rooms. A number significant differences were found, most often in tempo. Some of these were relatively easy to explain by examining time-series plots of the features, separated by room. For example, flutist 1 tended to play slower in the Salon des Nobles (see section 5.4.2.1), aligning with some previous findings that musicians tend to play slower in more reverberant rooms. However, some differences, though highly significant, were more difficult to interpret through the time-series plots.

It is important to note that any measurable differences in playing style between the two rooms, may not be solely due to the difference in acoustics between the two spaces. There are a number of factors which can influence a musical performance which are almost impossible to control for (see section 2.1). The experimental design tried to account for this in order reduce some of these unwanted effects, such as by having multiple musicians per instrument, and by randomizing the order in which the musicians experienced the rooms. However, it is still possible that some external factors may have had a systematic influence on the musicians' which was comparable to that of the acoustics. For example, theorbist 1 may have been influenced by the playing style of another musician whom they only encountered in one space as indicated through

contemporaneous observation (see section 5.4.3.1). Additionally, it is difficult to separate the acoustic attributes of the Salon des Nobles from its visually impressive and historical nature. In fact, several musicians mentioned in their responses to open-ended questions that they felt inspired by the visual ornamentation in the Salon des Nobles. It is therefore possible that any systematically measured differences in playing style could be due in part to one of these other factors.

Despite this, some responses to the open-ended questions provided evidence that the difference in acoustics between the two rooms was noticed by the musicians and caused them to respond in similar ways. For example, all of the flutists, and some of the other musicians, stated that they attempted to elongate the duration of notes as a way to compensate for the lack of resonance or reverberation in the amphitheater. In some cases this is observable in the objective measurement data as in sections 5.4.2.1 and 5.4.2.3.

These performance changes were part of an adjustment strategy that has been previously observed in other studies (see section 2.3). Furthermore, these performance changes are meaningful from a musicological point of view in that the elongation of notes, in response to the lack of reverberance in the amphitheater, results in a more legato and therefore less baroque-appropriate performance style. In other words, the acoustics of the amphitheater compelled some musicians to play in a less baroque-appropriate way, whereas this pressure was not evident in the Salon des Nobles. In contrast, most musicians had strong positive feelings about the acoustics of the Salon des Nobles and stated that it supported their playing.

These changes were more evident among flutists than either violists or theorbists. This may be due to the fact that both theorbists and violists have larger instruments with resonating bodies that create their own acoustical decay (especially compared to flutes), which may lessen the relative impact of the room acoustics. Flutists, on the other hand, may rely more heavily on the room acoustics to support the sound of their instrument and therefore may be more susceptible to acoustic changes.

With only two rooms, it is impossible to observe trends or to make generalizations about these playing style differences as a function of acoustics. Without responses to the questionnaires, it would be nearly impossible to know whether any of these measured changes are the result of an expressive intention by the performer in response to the acoustics rather than the result of normal variation from one performance to another. The questionnaire responses in this case were crucial in providing additional context which proved to be essential.

One significant drawback to this method of music performance analysis is that many of the underlying features lack a direct musical meaning. For example, while tempo and dynamics

are important to executing an expressive music performance, it can be challenging to discern much about the performance's expressive qualities from the observed measures. An alternative approach is to develop a framework informed by musicological principles to identify musical features which are important to the specific performance style. The following chapter (chapter 6) delves into a distinct approach for analyzing music performance, specifically targeting key areas crucial to executing a historical baroque performance.

Lastly, although some significant differences were measured in tempo and intensity, these data do not reveal whether or not these differences are noticeable to a listener. In order to better understand the perceptibility of these changes, a listening test was performed which is detailed in chapter 7.

Chapter 6

Baroque analysis framework

The second main approach taken to analyze the music performances from chapter 4 was to design a musicologically-informed analysis framework specifically for identifying key musical parameters which are essential to baroque historically informed performance (HIP). This analysis framework was first verified on professional recordings representing a variety of baroque performance styles then applied to the recordings from chapter 4.

6.1 Justification

There have been many strategies for analyzing music performances, some of which have been reviewed in section 2.1.1. One common strategy is through the extraction of a high number of low-level features. While this can be an effective way of identifying differences in performances, interpreting these differences in a musically meaningful way is difficult. Furthermore, an indifferent approach towards feature selection can yield a noisy feature set full of redundancies. One way to address this is through careful selection of low-level features. However, the most commonly used low-level features may be limiting and often lack a direct musical meaning. Furthermore, it is not always known which features will be most discriminative, and one runs the risk of leaving out meaningful features whose capabilities may not be well known. Aside from discriminatively selecting low-level features, dimensionality reduction techniques are also

commonly used, as was done in chapter 5. Although this ensures a smaller feature set while retaining most of the desired variance, the resulting dimensions may be even more difficult to interpret than the initial low-level features.

An alternative strategy is to design a feature set, guided by musicological principles, which is appropriate for the task at hand. A primary goal of this strategy is to end up with fewer but more semantically meaningful features. Because each genre may have a different set of intrinsic mechanisms which contribute to an idealized performance within that genre, it is important that the analysis framework take into consideration the musicological context. Of course, this is a challenging endeavor and the efficacy of the designed features is difficult to verify without a large corpus of annotated data. Nevertheless, this chapter describes some first steps towards this experimental approach.

This custom analysis framework is first verified on a small set of professional recordings which musicologists and listeners have already identified as representing distinct baroque performance styles (see section 6.3.1). The validated custom features are then applied to the recordings from chapter 4 to examine whether the room has any measurable effect on performance in terms of these bespoke dimensions (see section 6.4).

6.2 Background

Musicologists have adopted the term *stylishness* as a way to describe musical expressiveness in terms of its appropriateness for a specific musical and historical context (Schubert and Fabian, 2006). For example, the parameters which define an expressive performance of 19^{th} century romantic music are very different from those which define an equally expressive performance of 18^{th} century baroque music. (Vibrato is a good example, which, in solo romantic is often used in abundance, whereas in solo baroque music it is used sparingly.) While both performances may be perceived as equally *stylish* their perceived *expressiveness* may vary widely due to this term's strong association with mainstream romantic musical gestures. For this reason, researchers have also adopted the term *baroque expressive* in order to differentiate a type of expressive performance which is deemed as being baroque stylish (Fabian and Schubert, 2009.)

While the historical baroque performance style has been detailed in section 2.1.3, some essential qualities will be reiterated here. Donington (1982, p. 167) states that "a transparent sonority, and an incisive articulation" are essential. Fabian and Schubert (2009) mentions "well defined metric groups" and "selectively used vibrato" as well as "the uneven bow strokes, the variety of tonguing patterns and their effect on tone qualities" as particularly important in

historical baroque playing style. While some of these performance characteristics may seem relatively subtle, listeners with only a basic music education have been shown to be capable of differentiating a baroque-expressive performance from other baroque playing styles (see section 6.3.1).

6.3 Feature design and verification

This section discusses the design and development of the analysis framework as well as the validation process using professional recordings representing distinct performance styles.

6.3.1 Dataset

Fabian (2003) reviewed and characterized a large set of recordings of solo violin music by J.S. Bach. These recordings were reexamined in Fabian and Schubert (2009), where the researchers identified three recordings which represented three distinct schools of baroque performance practice in a listening test. One of these was the historically appropriate *baroque-expressive* style previously discussed (see section 6.2). The other two performance styles are referred to as *expressive-emotional* and *modern-literalistic*. The expressive-emotional style is characterized by legato phrasing and intense vibrato. This approach is representative of the mainstream approach to baroque performance at the time of the recordings (the 1930s). The modern-literalistic style is characterized by a straightforward, unvarying approach to phrasing while mostly refraining from traditional expressive gestures. This recording is representative of a common approach to baroque performance of the 1960s, which has since fallen out of fashion.

In Fabian and Schubert (2009), a listening test was performed which found that listeners were able to identify these three schools of performance as evidenced by ratings within several aesthetic categories. Significant differences were found in three major areas. The first was *phrasing*, where the baroque-expressive performance was rated as less "continuous" and more "articulated." The second was in *tone production* where the baroque-expressive tone was judged to be lighter compared to the other two styles. And finally, the third was *vibrato* where the baroque-expressive performance was found to exhibit less vibrato than the other two performance styles.

The listening test in that study relied on an eight bar excerpt of J.S. Bach's *Sarabanda* from the D minor Partita for solo violin. However, relying only on this musical example to verify the proposed analysis framework would be inadequate, as it would not produce enough data to yield meaningful results. To address this, the number of recordings was expanded to

Table 6.1 Information on professional recordings used.

Playing style	Artist	Year	Label & No.
Expressive-emotional	Yehudi Menuhin	1934	Philips 438 736-2
Modern-literalistic	Arthur Grumiaux	1967	Nonesuch 7559 73030-2
Baroque-expressive	Sergiù Luca	1977	EMI CHS 7 63035 2

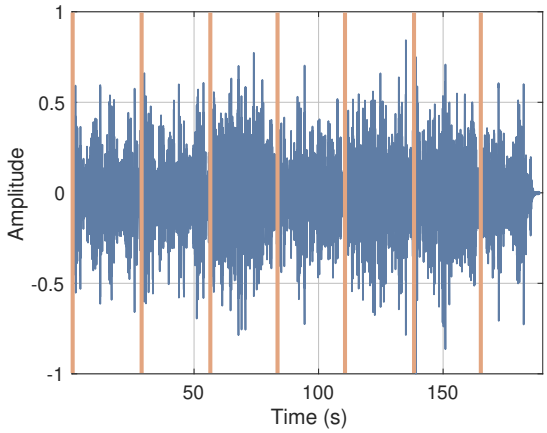
three full solo violin pieces by the same musicians and representing the three aforementioned performance styles. These additional pieces, all composed for solo violin by J.S. Bach, were the previously mentioned Sarabanda, the Gavotte en Rondeau from the E major Partita, and the Largo from the C major Sonata. These pieces were chosen to represent a larger variety of compositional styles within the baroque era. While these additional recordings have not been verified by listeners as representing specific performance styles, it is assumed that these additional recordings by the same musicians, from the same recording set, would carry similar performance characteristics to the one previously examined in that study.

The following sections detail the development of features to differentiate between baroque performance styles in the three major areas previously identified by listeners, namely, phrasing, tone production, and vibrato. The verification of each feature is performed by training support vector machines (SVMs) using a radial basis kernel and a one-versus-one approach with cross validation (5 random folds) to predict the musician, who serves as a proxy for performance style.

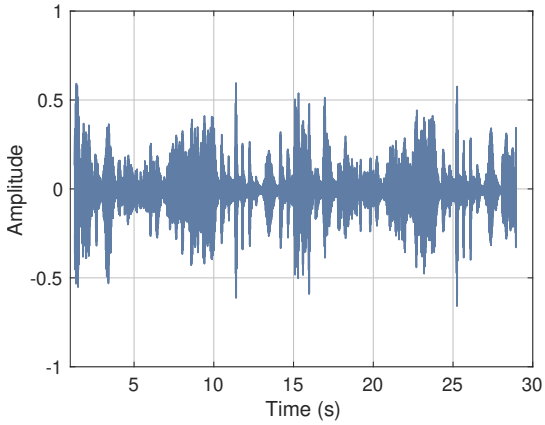
6.3.2 Phrasing

The approach to phrasing in historical baroque performance typically includes "well defined metric groups" rather than large phrases connected by legato and continuous playing (Fabian and Schubert, 2009). Additionally, this phrasing style has been described as *rhetorical*, since it tends to mimic the patterns found in oration, in that it is varied, locally nuanced, and rhythmically flexible (Ponsford, 2012). This stands in contrast to mainstream performance styles which tend to use continuous phrasing to highlight cadences and large phrase structures.The features developed to capture expressive phrasing must therefore be capable of simultaneously highlighting both local and global differences.

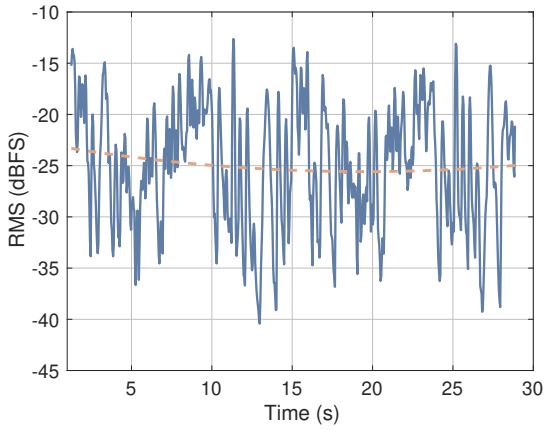
Previous research on expressive phrasing in music performance have found that musicians are largely able to accomplish it through manipulation of tempo and loudness (Palmer, 1989). Therefore, the proposed features aim to capture large-scale and small-scale variations in both



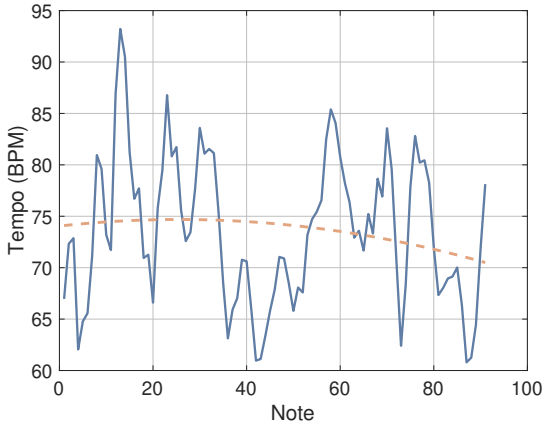
(a) Full performance with 16-bar segments marked.



(b) Excerpt of first 16-bar segment.



(c) Loudness curve of first 16-bar segment from which the range, standard deviation, and coefficients of a 2nd-order polynomial (overlaid) are computed.



(d) Tempo curve of first 16-bar segment from which the range, standard deviation, and coefficients of a 2nd-order polynomial (overlaid) are computed.

Figure 6.1 Procedure for calculating phrasing features. This example is from the baroque-expressive recording of J.S. Bach's *Gavotte*.

of these parameters.

Since the exact length of all phrases is not known ahead of time, the recordings were separated into segments of different generic lengths (1, 2, 4, 8, and 16 bars) and features were extracted from the tempo and loudness curves of these segments, as well as the entire recordings. While this may fail to capture all phrase boundaries, some of which may fall at the end of odd-numbered measures, it was chosen over an approach such as in Chuan and Chew

(2007) which aims to identify phrase boundaries through more advanced modeling techniques, due to its simplicity.

The tempo curves were created based on note-level tempo values which were smoothed with a moving average filter (with a span of 3). Smoothed tempo curves have been previously used to successfully model expressive timing (Cambouropoulos et al., 2001; Schreiber et al., 2020). Furthermore, absolute representations of tempo were found to be more correlated with human perception as compared to normalized tempo representations (Timmers, 2005). The loudness curves were created by calculating the root mean square (RMS) amplitude on frames of length 100 ms with a hop size of 0.04 s.

From these curves, statistical descriptors were calculated including the range, standard deviation, and coefficients of a parabolic function (2^{nd}-order polynomial curve). These statistical descriptors were arrived at through a combination of a literature review, which found that previous studies have shown success in using polynomial coefficients to model expressive performances (Li et al., 2017), and trial and error. The precise polynomial order was chosen since it has been shown to be effective in many previous studies (Todd, 1992; Repp, 1998, 1999a; Timmers, 2005). Overall, this collection of violin recordings yielded 1029 segments which consisted of 98, 206, and 39 segments for the Sarabanda, Gavotte en Rondeau, and Largo, respectively, for each playing style.

These phrasing features were used to train a SVM classifier tasked with predicting the musician (representing performance style), the classifier achieved an average F-score (the harmonic mean of precision and recall) of 72.5%. The full precision and recall for the SVMs used to predict the musician can be found in table 6.2.

As further exploration, the same classifier trained on the same features was tasked with predicting the composition. This was thought to be an easier task than predicting performance style, since the compositions more clearly represent differences in these dimensions compared to more subtle phrasing differences attributed to performance style. The results, therefore, should suggest an approximate upper bound of the discriminability of these features. The resulting average F-score was 87.6% suggesting these features are able to discern differences in style quite well. A visualization in the form of a t-dstributed stochastic neighbor embedding (t-SNE) projection intended to demonstrate the discriminability of these features in differentiating between both sets of classes can be seen in fig. 6.2).

Table 6.2 Precision, recall, and F-scores for a SVM classifier trained on phrasing features of Bach violin recordings to predict musician which serves as a proxy for performance style.

Class	Precision	Recall	F-Score
Expressive-emotional	71.7%	78.2%	74.8%
Modern-literalistic	70.4%	68.5%	69.4%
Baroque-expressive	75.7%	70.8%	73.2%
Average	72.6%	72.5%	**72.5%**

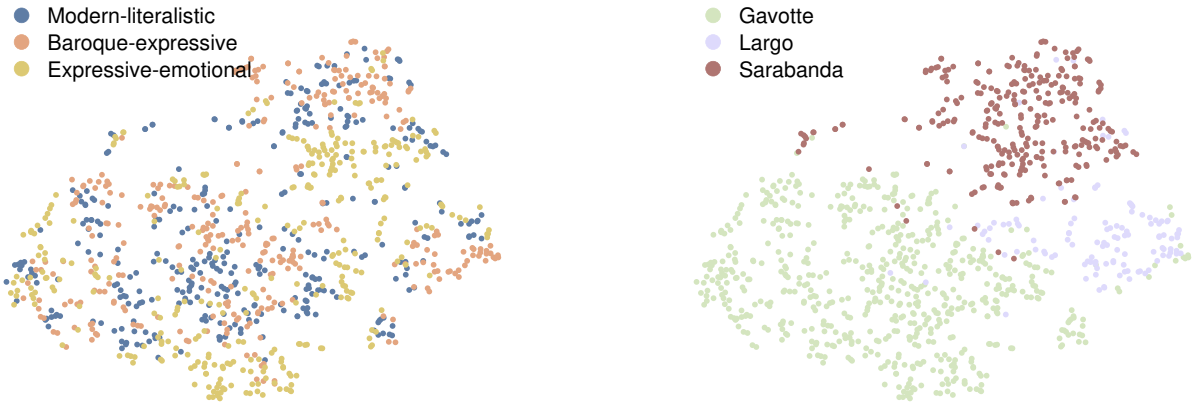


Figure 6.2 t-SNE projection of the phrasing features showing discrimination between (a) performance style and (b) composition.

6.3.3 Tone production

One aspect of tone production which is an important part of baroque expressiveness, and which listeners were able to identify in Fabian and Schubert (2009), is "lightness of tone." The listeners rated the baroque-expressive performance as having exhibited a lighter tone compared to the other two performance styles. The notion of lightness of tone is difficult to define succinctly, Donington (1982, p. 167) is certainly alluding to it when he describes "a transparent sonority, and an incisive articulation" as being essential to baroque HIP. It can also be associated with articulation technique, where a lighter tone may be the result of less bow pressure among string players or less air pressure among wind players (Schubert and Fabian, 2014).

It is difficult to separate the aspect of tone production, which results from an intentional action by the musician, from that which is a result of the design and fabrication of the instrument

and/or bow. This is especially true when the difference between instruments is very distinct, as in the case of a modern versus baroque violin. Although, Donington (1982, p. 165) implies that the instrument's tonal identity ultimately has an influence on the musician's actions and that these two domains are not as distinct as they appear, claiming, "the *sound* of baroque music can only be recovered on its own instruments... the *style* is very largely dependent upon the sound." In other words, the sonority of the instrument necessitates the musician to adapt a particular playing style which produces a tone which is more amenable to historical baroque style. These two components become self-reinforcing.

Tone production is certainly correlated with timbre, especially how timbre evolves over time. Timbre itself, however, is multidimensional and complex. It includes subjective attributes such as "color", "shape", and "texture" and may be correlated with loudness, pitch, and duration (McAdams, 2013). Timbre can help inform a listener of a sound object's identity and also its quality. The latter attribute is more of interest in this case. Identifying this somewhat ambiguous attribute in a systematic way is very challenging, however, identifying differences in timbre, which may suggest a significant difference in tone production is possible.

Capturing qualitative aspects of a sound, such as tone production, through signal processing techniques is a known challenge (Knight et al., 2011). In fact, most experiments seeking to elucidate tone quality rely on subjective tests using dissimilarity ratings (McAdams et al., 1992). Signal processing approaches are often used for classification tasks, such as instrument identification (Siedenburg et al., 2016; Fragoulis et al., 2006). One major challenge to signal processing approaches to timbre is that it is difficult to capture the desired attributes of a sound without also capturing other unwanted aspects present in the signal (such as the quality of the recording, the acoustics of the recording space, or the specific instrument used).

One of the most common approaches in music information retrieval (MIR) for extracting timbre information from a signal is through mel-frequency cepstral coefficients (MFCCs). One reason these features are commonly used is because they model, to some extent, certain attributes of the human auditory system such as a non-linear frequency resolution and compression. For this reason, they are believed to have a strong perceptual basis, although there is some disagreement on how much this assumption can be relied upon (Aucouturier and Bigand, 2012). Regardless, they have shown strong results in many audio classification tasks (Siedenburg et al., 2016; Fragoulis et al., 2006).

Despite some known disadvantages, MFCCs were used in this chapter to try to capture information about tone production. A window size of 8192 samples and a hop size of 4096 samples were used and 13 coefficients were employed. In addition to extracting MFCCs from the

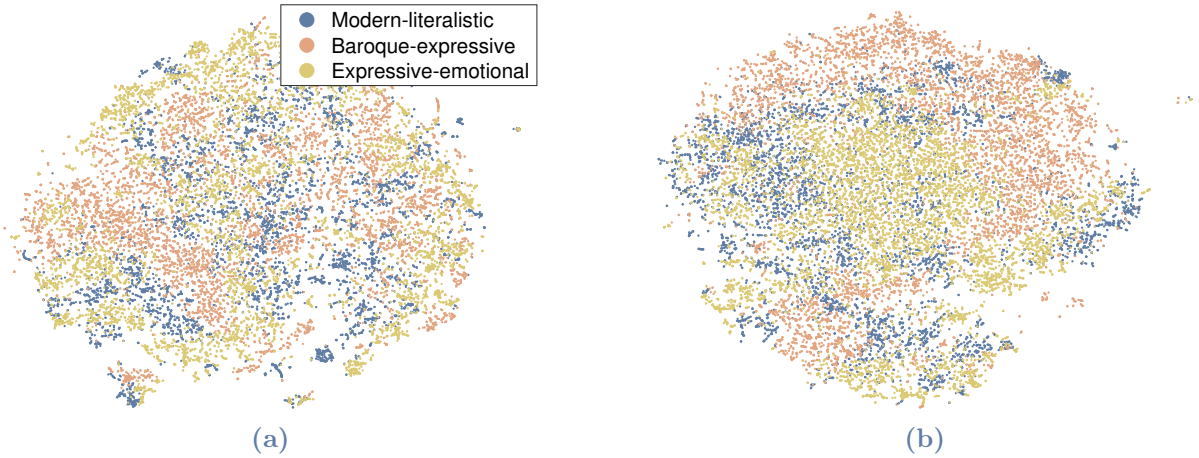


Figure 6.3 t-SNE projection of MFCCs derived from the (a) unprocessed audio and the (b) percussive audio showing a slightly better class separability for the baroque-expressive class using the percussive components.

recordings, they were also extracted from the derived harmonic and percussive components of the recordings. The motivation for this approach was to see if either of these components could reveal more essential aspects about the tone production that might be obscured in the original signal. Of particular interest was the percussive component which emphasizes the nontonal, transient, and stochastic elements of the signal. These elements are likely correlated with some essential aspects behind an instruments articulatory mechanisms, such as bowing, which is important to one's perception of sound quality (Schoonderwaldt et al., 2003). Furthermore, the nontonal spectrum has previously been used with success to identify different instruments in classification tasks (Fragoulis et al., 2006).

The audio recordings were separated into their harmonic and percussive components using the median filter approach (Fitzgerald, 2010). This approach applies a median filter to the short-time fourier transform (STFT), or spectrogram, either across successive bins or frames, which enhances either the harmonic or percussive components of the signal while suppressing the other component. A typical use case for this kind of harmonic and percussive separation is to separate the drums from a pop music recording to facilitate remixing.

The original, harmonic, and percussive sets of MFCCs were used to train SVM classifiers using the same hyperparameters as previously described, which were tasked with predicting the musician (performance style). The class-weighted average F-scores are reported in table 6.3. Overall, the MFCCs derived from the unprocessed audio performed the best, with an F-score of

Table 6.3 F-scores for SVM classifiers trained on tone production features (MFCCs derived from unprocessed audio as well as its harmonic and percussive components) of Bach violin recordings to predict performance style.

	Unprocessed	Harmonic	Percussive
Expressive-emotional	93.2%	90.8%	87.5%
Modern-literalistic	91.4%	86.0%	83.9%
Baroque-expressive	94.4%	88.6%	91.4%
Average	**93.0%**	88.6%	87.5%

93.0%. The other two sets of MFCCs were slightly less effective, but produced similar results to each other, with the harmonic-derived MFCCs resulting in an F-score of 88.6% and the percussive-derived MFCCs resulting in an F-score of 87.5%.

Despite the average F-score of the percussive-derived MFCCs being lower than the unprocessed ones, the F-score of the baroque-expressive class is significantly higher than the other classes when trained using MFCCs derived from the percussive components. Observing the t-SNE projections of the MFCCs from both the unprocessed and percussive audio (seen in fig. 6.3), the baroque-expressive class shows slightly more separation in the percussive set. This suggests that the percussive components of the audio may be highlighting certain parts of the signal which are important in identifying a baroque-expressive tone quality. However, a larger, more representative dataset would be needed to verify this.

It is still unknown precisely how much this class discriminability is due to parts of the signal which are unrelated to tone production, such as the acoustics of the recording space, or other aspects about the recording setup (microphone choice, equalization, etc.). This concern is explored in more detail in section 6.5, specifically in regard to quantifying the influence of the room acoustics on these features.

6.3.4 Vibrato

Vibrato is a musical ornamentation whereby semi-periodic fluctuations of pitch (and often loudness) are used to embellish a sustained tone. It is often described by two features: *rate* and *extent* (Seashore, 1938). Rate refers to the frequency of modulation which is typically in the range of 4 Hz to 8 Hz (Weninger et al., 2012). Extent refers to the depth or intensity of the

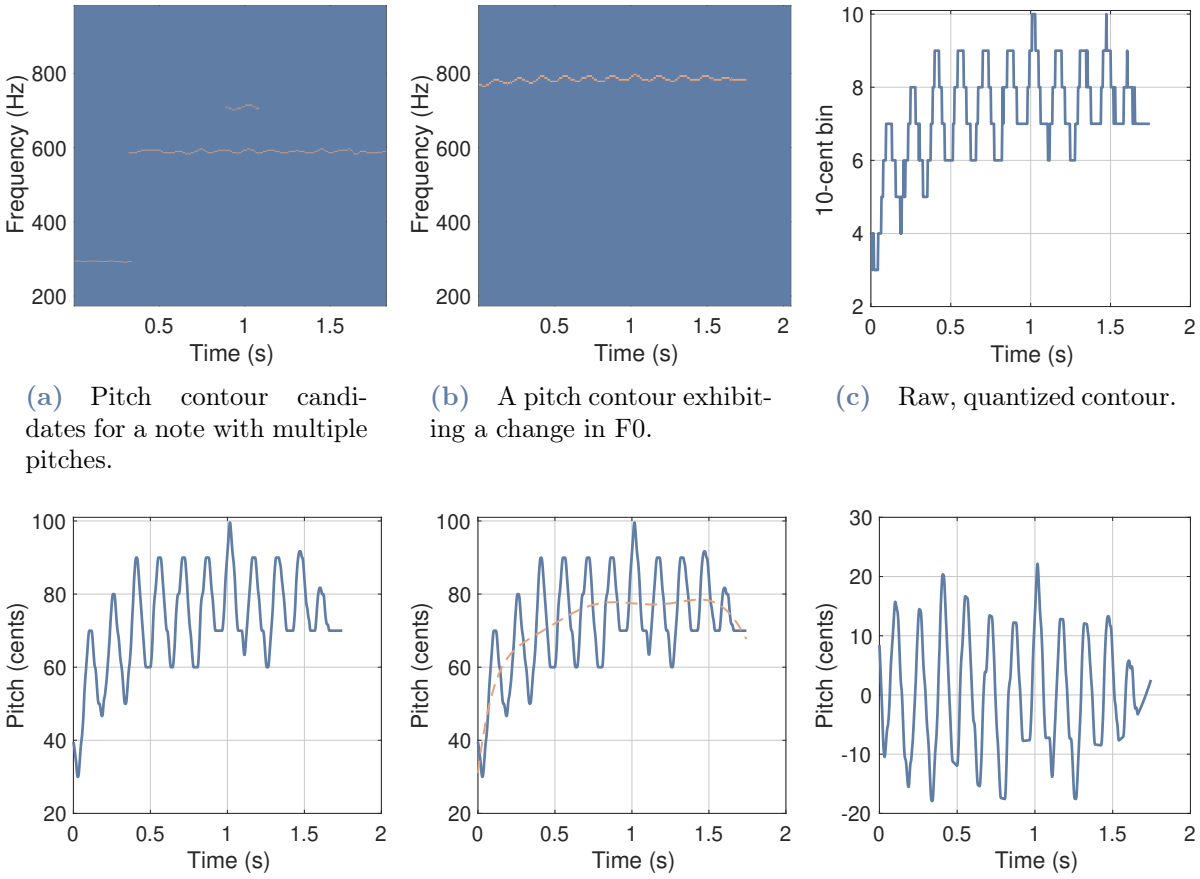
vibrato, measured in cents.

Vibrato is commonly achieved through a combination of pitch and loudness modulation. However, instruments that lack the capability to make small pitch adjustments may resort to modifying only the loudness to embellish a sustained note, a technique commonly known as tremolo. For the purposes of this section, only the traditional interpretation of vibrato, where the pitch is varied in a semi-periodic manner over the duration of a sustained note, will be considered. Baroque-expressive performances typically use vibrato sparingly and with less intensity, or a narrower extent than other performance styles, including the previously mentioned modern-literalistic and expressive-emotional styles (Fabian and Schubert, 2009).

Most published methods calculate extent in reference to a pitch center, usually the mean (Herrera and Bonada, 1998; Weninger et al., 2012; Pang and Yoon, 2005; Regnier and Peeters, 2009). However, this assumes that the fundamental frequency remains unchanged throughout the duration of the note. While this may be the case in controlled laboratory experiments such as in Brown and Vaughn (1996), in real performances the pitch center of the note may vary throughout the duration of the note while maintaining vibrato. A study that was carried out as part of this thesis applied linear regression to address the issue of a varying pitch center when calculating vibrato parameters. That study investigated the impact of room acoustics on singers of medieval music. The development of the analysis tools in that study facilitated the development of the methodology for calculating vibrato for this baroque analysis framework. However, because that study is not directly related to the main objectives of this thesis it will not be thoroughly discussed here. The complete article can be located in appendix B.

The proposed methodology for calculating vibrato features attempts to improve upon previous published methods in two major ways. First, the pitch center of the note is modeled in a variable way, using either the mean or with a polynomial fit curve. The method is chosen based on the kurtosis of the segment of the signal of a note which has been identified as exhibiting vibrato (which serves as an estimation of the pitch stability over the duration of the note) with an empirically derived threshold. This is intended to improve the accuracy of the measured extent of the vibrato. Second, the evolution of the vibrato rate throughout the duration of the note is modeled using polynomial fit coefficients rather than relying on a single metric (such as the mean or median) to describe the vibrato rate.

Rather than using a fundamental frequency estimation algorithm such as pYin (Mauch and Dixon, 2014) for identifying vibrato, pitch contours (Salamon and Gomez, 2012) were used. Pitch contours have traditionally been used for melody extraction from polyphonic signals. They are continuous curves of varying lengths that can output multiple pitch candidates at a



(a) Pitch contour candidates for a note with multiple pitches.

(b) A pitch contour exhibiting a change in F0.

(c) Raw, quantized contour.

(d) Upsampled and smoothed contour.

(e) Smoothed contour with overlaid estimated pitch center.

(f) Contour with pitch center removed.

Figure 6.4 Vibrato feature extraction procedure.

given time, which gives them an advantage over fundamental frequency estimators in that they are more robust in polyphonic settings or for instruments which can play multiple tones at once, such as string instruments. Pitch contours have also been shown to be discriminative as mid-level features including for characterizing vocal style from derived vibrato characteristics (Bittner et al., 2017).

The pitch contours were calculated using the the Melodia plug-in[1] for Sonic Visualiser. They were then segmented by note, using note onset information which was calculated and manually verified as described in section 5.1.1. In cases where multiple pitch candidates are

[1] https://www.upf.edu/web/mtg/melodia

Table 6.4 Class-weighted average F-scores for SVM classifiers trained on vibrato features extracted from Bach violin recordings to predict performance style. The mean and standard deviation of 100 classifier runs are reported.

Features	Pitch center	F-Score (mean ± SD)
Baseline	Mean	45.1% ± 0.8
Baseline	Variable	48.9% ± 0.5
All	Mean	44.9% ± 1.0
All	Variable	**50.0**% ± 0.6

detected within the span of a single note (as in fig. 6.4a), the most salient contour was chosen by summing the energy in each bin (representing 10 cents), then selecting the bin with the most energy. This bin, in addition to the six adjacent bins above and below (equivalent to a range of 130 cents) are then isolated for further processing (see fig. 6.4c).

The contour was then upsampled, smoothed using a moving average filter (with a span of 100), and factored by a gain of 10 so that each bin would represent 1 cent (see fig. 6.4d) in order to facilitate further processing. The note was considered as exhibiting vibrato if the signal met three threshold conditions: a minimum length, zero crossing rate (on the pitch-centered signal), and extent. These thresholds were determined empirically through preliminary testing.

If the presence of vibrato was indicated, the pitch center was approximated using either the mean or a polynomial curve of degree 5 (see fig. 6.4e). The pitch center was then subtracted from each value, resulting in the final signal (seen in fig. 6.4f) from which the vibrato parameters (rate and extent) were calculated.

The extent was calculated by taking the mean of the doubled absolute value of each peak and trough. The rate was calculated frame-wise, taking an auto-correlation of the signal followed by peak picking, a basic fundamental frequency estimation method (Lerch, 2012, pp. 98–99). A peak distance threshold was selected so that the output would be bound within the typical expected rates of vibrato in order to avoid octave errors. The median value of all frames was selected as the note-level rate. Polynomial fit coefficients of order 3 were taken of the frame-wise rate in order to model the change in rate throughout the duration of the note.

To compare different vibrato features and extraction methodologies, four classification tasks were considered. Half of the tasks used the common method of estimating the pitch center using the mean, while the other half used a variable estimation method (either using a polynomial curve or the mean to estimate the pitch center). For each of the pitch center estimation

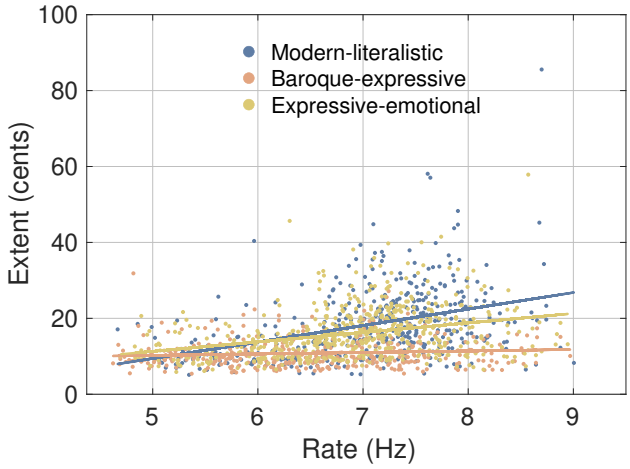


Figure 6.5 A scatter plot of vibrato parameters rate and extent, separated by performance style. Linear regressions for each class show the correlation (or lack thereof) of the two parameters as a function of performance style.

methodologies, there were two different feature sets: the baseline feature set which included only the rate and extent, and the extended feature set which included the baseline features in addition to polynomial fit coefficients intended to model the change in vibrato rate over the duration of the note. Overall, 972 notes with vibrato were identified out of 4374 played notes.

As with the previous features, the vibrato features were used to train SVM classifiers tasked with predicting the musician. Several SVMs were trained on vibrato features derived from different methodologies and feature sets. Because the various methodologies were yielding fairly similar results, and because the random folds of the classifier can yield slightly different results every time, the SVMs were trained 100 times. The mean and standard deviation of the resulting F-scores of those trials are reported in table 6.4.

The results show that the proposed improvements produced slightly better classification accuracy compared to previous methodologies. These F-scores are all significantly above random chance (33.3%), suggesting modest success in separating the performance style based on vibrato features alone. The proposed methodological changes and additional features yielded a 5% increase in F-score.

Various vibrato characteristics (shown in table 6.5) reveal that the baroque-expressive style exhibits the least amount of vibrato, in terms of prevalence (the percentage of total notes that exhibit vibrato), rate, and extent. This aligns with musicological expectations. Furthermore, the expressive-emotional style shows the most prevalent vibrato, as expected. Both the

Table 6.5 Vibrato characteristics from Bach violin recordings for each performance style.

Style	Prevalence	Rate (mean ± SD)	Extent (mean ± SD)
Expressive-emotional	32%	7.1 Hz ±0.7	17.4 cents ±6.5
Modern-literalistic	22%	7.3 Hz ±0.7	20.3 cents ±8.3
Baroque-expressive	13%	6.8 Hz ±0.9	12.9 cents ±4.0

expressive-emotional style and the modern-literalistic style show somewhat comparable rates and extents, however.

Figure 6.5 shows a scatter plot comparing the rate and extent, separated by performance style. One can see that the rate and extent of the baroque-expressive style tend to be lower than the other performance styles. Another difference in these performance styles was found in the relationship between these two variables. A linear regression line is overlaid on the data to demonstrate that there is a moderate correlation between these two variables within the modern-literalistic ($r = 0.40$, $p < .001$) and expressive-emotional ($r = 0.32$, $p < .001$) performance styles. By contrast, the baroque-expressive style does not demonstrate a significant correlation between these two variables ($r = 0.10$, $p > .05$). This reveals another dimension by which vibrato differs as a function of the examined performance styles. These data were also fit with polynomials of various degrees but these did not yield significantly better fits.

It can be assumed that the vibrato prevalence has a strong influence on a listener's perceived amount of vibrato, probably more than the rate and extent. The classification tasks did not take these global values into account. A more sophisticated multi-level model taking this into account could improve the characterization of vibrato styles.

6.4 Applied analysis

The features developed and validated in the previous section were applied to the recordings from the two real rooms (the Salon des Nobles and the amphitheater) described in chapter 4. The goal was to gain a better understanding of how the playing style differed between the two rooms within these dimensions regarded to be important to baroque expressiveness.

6.4.1 Applied phrasing analysis

SVMs were trained on the phrasing features extracted from these recordings and tasked with predicting the room. The classification framework used in chapter 5 was used again here, where the classification of each performance is based on a majority vote of the classification of each of its constituent observations. However, when classifying all pieces together, two approaches were taken, one using the majority vote method described above (denoted with the subscript $_{maj}$) and one in which each observation was classified individually (denoted with the subscript $_{ind}$). The purpose of taking these two approaches is to facilitate comparisons with other approaches. For example, the $_{maj}$ classification allows straightforward comparisons to the results in chapter 5, while the $_{ind}$ results are more easily comparable to the results in sections 6.3 and 6.5. The classification schemes were limited to those that included all musicians since the number of observations for performance was quite low compared to the analysis in chapter 5.

The results, shown in the top half of table 6.6, demonstrated perfect classification when using both the tempo and loudness features for all instruments using the majority vote classification method. This means that every performance had a majority of observations classified as having taken place in the correct room. Compared to the results in section 5.3 using the tempo and loudness features (table 5.3), which had an accuracy about 67% within each instrument class when trained on all pieces, this is a major improvement.

As before, there is a concern that some of the loudness features may be influenced by the sound of the room present in the recordings (see section 6.5 for more information on this topic). For this reason, this classification task was repeated using SVMs which had been trained only on the tempo-related phrasing features.

The results using only tempo-related phrasing features, (bottom half of table 6.6) exhibit fairly high accuracy, with most individual pieces exhibiting an accuracy above 80%. Each classification scheme outperforms or is equal to the corresponding classification scheme reported in table 5.4 from section 5.3, with accuracy improvements generally greater than 20%. Notably, the accuracy improvements for the viol and flute are much more significant than those for the theorbo which showed mostly mediocre improvements. With the previously discussed knowledge that the theorbists had some difficulty performing without error, this is not a surprise.

Using these proposed phrasing features, considerable improvements in classification accuracy are seen over previous efforts in chapter 5 (see tables 5.3 and 5.4) using what are essentially the same elements of tempo and intensity. Using this method, the classifier obtained 100% accuracy (using the majority vote method) for all instruments, while the tempo and loudness features from chapter 5 only yielded an accuracy of about 67% for each instrument. Restrict-

Table 6.6 Classification results for phrasing features applied to recordings from study in chapter 4. The $_{maj}$ and $_{ind}$ subscripts denote different methods of classification where $_{maj}$ is through a majority vote of all observations in a single performance and $_{ind}$ denotes each observation classified individually.

Feature set	Viol		Flute		Theorbo	
Tempo and loudness	Couplet 12	100.0%	Animé	100.0%	Gigue A	100.0%
	Couplet 13	100.0%	Lentement	100.0%	Gigue B	100.0%
	Couplet 18	100.0%	Gravement	100.0%	Prelude	100.0%
	All_{maj}	100.0%	All_{maj}	100.0%	All_{maj}	100.0%
	All_{ind}	89.5%	All_{ind}	95.0%	All_{ind}	86.9%
Tempo only	Couplet 12	95.8%	Animé	94.7%	Gigue A	78.1%
	Couplet 13	87.5%	Lentement	100.0%	Gigue B	79.0%
	Couplet 18	95.8%	Gravement	100.0%	Prelude	79.0%
	All_{maj}	83.3%	All_{maj}	84.2%	All_{maj}	68.6%
	All_{ind}	63.4%	All_{ind}	73.4%	All_{ind}	60.7%

ing these to only tempo-related features, this method yielded viol and flute accuracies of 83.3% and 84.2%, respectively, whereas the tempo features from chapter 5 resulted in viol and flute accuracies of 69.4% and 61.4%, respectively. The classification accuracy for the theorbo performances was not greatly changed between the two methods using only tempo features, although this is likely due to reasons previously discussed (such as the difficulty the theorbists had in performing the repertoire without mistakes).

This indicates that the process of using statistical descriptors to characterize musically meaningful segments is able to highlight latent differences within the performances that simple synchronized time series data (as were used in chapter 5) was not. It is likely that this process results in less noisy data with fewer but more meaningful observations.

6.4.2 Applied tone production analysis

The approach to tone production described in section 6.3.3 was applied to the recordings from the main experiment described in chapter 4. As before, SVMs were trained to predict the room. Each of the observations was classified separately and the overall accuracy is shown in table 6.7.

The results show that the SVMs trained on these features were able to predict the room

Table 6.7 Overall accuracy of SVM classifiers trained on tone production features from recordings from the main experiment to predict room.

	Unprocessed	Harmonic	Percussive
Viol	**88.5%**	77.5%	77.4%
Flute	84.2%	**86.1%**	81.5%
Theorbo	**74.3%**	61.5%	64.3%

with fairly high accuracy, typically over 80% for the MFCCs derived from the unprocessed audio. There is evidence that many of the musicians adjusted their tone production due to the acoustics, especially as a strategy to adapt to the less reverberant amphitheater where some musicians felt compelled to project more (see section 5.4). However, it is difficult to tell to what extent these features are identifying these differences in tone production compared to other components of the signal, such as the presence of the room sound.

The viol and theorbo results show that the MFCCs derived from the unprocessed audio have the best predictive power, as was the case in the violin dataset (see section 6.3.3). However, this was not the case for the flute, where the MFCCs derived from the harmonic audio components gave the best results. Furthermore, the difference in accuracy among the different MFCCs is larger for the viol and theorbo ($> 10\%$), whereas it is a bit smaller for the flute ($< 5\%$).

The lower accuracy for the theorbo is not a surprise for reasons previously discussed. It is possible that the results for the theorbo may suggest an approximate lower bound for the power of these features to discriminate tone quality, as opposed to other elements present in the recordings.

6.4.3 Applied vibrato analysis

The analysis of vibrato was only performed on the viol because it is the instrument that can most clearly produce this technique. While the theorbo can also manipulate pitch to create vibrato, it requires significantly more effort than the viol and is therefore rarely used for this purpose. Additionally, flutists have a similar ornamentation technique to vibrato, in which the amplitude is varied in a periodic way, but this was not included in the analysis as it focused solely on detecting pitch manipulations.

In total, 1143 notes were identified as having vibrato out of 6888 played notes, translating to a prevalence of about 17%. Various vibrato characteristics are summarized in table 6.8. The overall prevalence among the entire test set of Bach solo violin recordings was about 22% while

Table 6.8 Vibrato characteristics of performances from main experiment separated by room.

Room	Prevalence	Rate (mean ± SD)	Extent (mean ± SD)
Salon des Nobles	18%	6.9 Hz ±0.9	15.3 cents ±6.3
Amphitheater	15%	6.8 Hz ±0.9	15.2 cents ±6.1

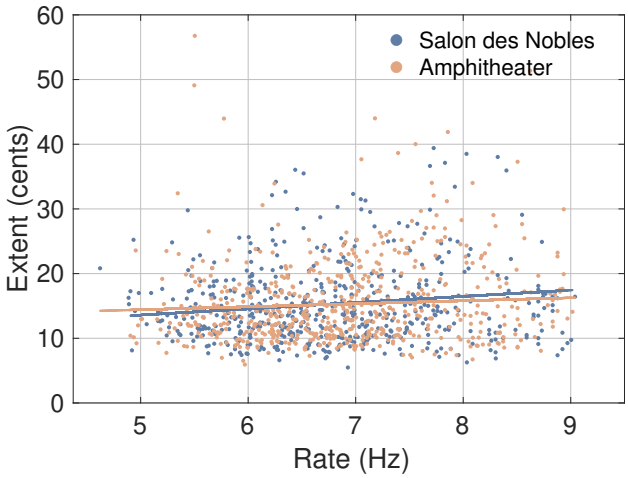


Figure 6.6 A scatter plot of vibrato parameters rate and extent, for recordings from the main experiment separated by room. Linear regressions show a lack of correlation between the two parameters.

the prevalence for the baroque-expressive Bach performances was 13%. The average vibrato rate was 6.8 Hz which is the same average rate among the baroque-expressive recordings from the test set. The average extent was 15.2 cents which is slightly higher than the baroque-expressive recordings from the test set (12.9 cents) but still lower than the average extent of the other two performance styles (17.4 cents and 20.3 cents for the expressive-emotional and modern-literalistic styles, respectively). While the Salon des Nobles has a slightly higher vibrato prevalence, meaning they are slightly less baroque-expressive in this dimension, the difference is not significant enough to merit further investigation.

In general, the difference in vibrato between the two rooms is not very significant. This was confirmed by a SVM classifier which was trained on the vibrato features, but performed no better than random chance. Furthermore, the scatter plot shown in fig. 6.6 provides further evidence for this. The correlation between the rate and extent is fairly low in both the Salon des Nobles ($r = 0.14$, $p < .001$) and the amphitheater ($r = 0.07$, $p > .05$), as shown by the

Table 6.9 Vibrato characteristics for each violist from the main experiment.

Musician	Prevalence	Rate (mean ± SD)	Extent (mean ± SD)
Violist 1	15%	6.9 Hz ±1.0	14.5 cents ±4.9
Violist 2	17%	6.9 Hz ±0.9	15.9 cents ±6.9
Violist 3	14%	6.7 Hz ±0.9	15.0 cents ±5.5
Violist 4	20%	6.8 Hz ±0.9	15.7 cents ±7.1

linear regression lines in fig. 6.6. Polynomials were also used to fit these data but did not yield better fits, suggesting very little to no correlation between vibrato rate and extent in these performances.

The vibrato characteristics for each violist are reported in table 6.9. It appears that each violist used vibrato with a fairly similar rate and extent. There is a slight difference in prevalence among the musicians but likely not enough to merit further investigation.

6.5 Influence of room sound on features

In examining the results for both methodologies of performance analysis, it has been difficult to assess the extent to which the sound of the room may have affected the metrics used to quantify the recorded performances. While the microphone choice and position were intended to maximize the direct-to-reverberant ratio without obstructing the musicians, some room sound is audible in the recordings. This section describes a small experiment that was designed to better quantify the extent to which the features used in this chapter may have been influenced by this incidental room sound. These findings can help contextualize the results by giving an indication of the lower bound of the performance of these features when applied to recordings from the main experiment.

6.5.1 Methodology

Anechoic recordings of flute and cello performances were convolved with room impulse responses (RIRs) from the two different calibrated geometrical acoustics (GA) models described in section 3.4. These GA models were adjusted to simulate the source-receiver setup during the actual experiments including a cardioid receiver. Two variations of these RIRs were exported: one with an omnidirectional source directivity and one with a trumpet source directivity. These two radiation patterns were chosen as they represent two relative extremes;

Table 6.10 F-scores of SVM classifiers trained on various features from anechoic recordings convolved with different RIRs to predict room (or which RIR the recording was convolved with).

Source Directivity	Phrasing		MFCC Norm.		MFCC Harm.		MFCC Perc.	
	Flute	Cello	Flute	Cello	Flute	Cello	Flute	Cello
Directional	74.5%	63.3%	45.6%	53.5%	63.9%	37.8%	40.3%	53.0%
Omnidirectional	79.4%	60.7%	57.9%	73.0%	72.5%	54.1%	55.6%	69.8%

the trumpet directivity represents a scenario which would have a greater direct-to-reverberant ratio, and therefore less incidental room sound, and the omnidirectional directivity represents the opposite extreme. The trumpet directivity examples will be referred to as "directional" in order to avoid confusion with the instruments used.

Three flute and two cello anechoic recordings were used. The flute recordings consisted of the first 16 measures of J.S. Bach's *Badinerie*, an excerpt of the first movement of W.A. Mozart's Flute Concerto in G, and an excerpt from Claude Debussy's *Syrinx*. The two cello recordings were 8 measures of the main melody from Gabriel Fauré's *Sicilienne* and 21 measures of the second movement of Franz Schubert's piano trio. The duration of the combined recordings is approximately 65 s for the flute and 70 s for the cello.

These recordings were convolved with the previously mentioned RIRs. A casual listening assessment confirmed that the direct-to-reverberant ratios in these recordings was comparable to the recordings from the actual experiment, especially those convolved with the RIRs using the directional directivity pattern.

The phrasing and tone production features were extracted from these recordings using the same methodology as previously described. A SVM classifier was trained to predict the "room," or, in other words, which RIR the recording was convolved with. A lower score would indicate that the feature is less influenced by the presence of the room sound. Because the number of observations in these examples was limited, two random folds were used in the cross-validation of the classifiers trained on the phrasing features and four random folds were used for those trained on the tone production features. The classifiers were run 1000 times, and the average accuracy from the 1000 trials is reported in table 6.10.

6.5.2 Results

The results are reported in table 6.10. A higher classification accuracy indicates that the underlying features are more sensitive to the room sound.

The results of the classifiers trained on the phrasing features suggest that an accuracy of up to around 79% may be due to differences in room sound alone. While the results using the omnidirectional RIR suggest a threshold of 60.7% for the cello and 79.4% for the flute, the results derived from the directional RIRs are probably closer to reality, suggesting that between 63.0% (cello) to 74.5% (flute) of the classifier performance is due to the room sound. While these are fairly high, the results from the recordings from the main experiment (see table 6.6) are over 20% higher (89.5% for the viol and 95.0% for the flute) suggesting that there are still meaningful differences between performances in the two rooms which were captured by these features.

The results of the classifiers trained on the tone production features suggest that these features are generally less affected by the presence of the room sound than the phrasing features. Classifiers trained on MFCCs based on the unprocessed, harmonic, and percussive audio yielded accuracies ranging from 37.8% (well below random chance, suggesting no influence from the room sound) to 73.0%. The highest accuracy using the omnidirectional RIR was from the unprocessed MFCCs of the cello (73.0%) while the highest accuracy from the directional RIR was from the MFCCs derived from the harmonic components of the flute recordings (63.9%). The best results from the tone production features in section 6.4.2 are 88.5% for the MFCCs from the unprocessed viol recordings (compared to 53.5% for the directional cello recordings here) and 86.1% for the harmonic MFCCs of the flute recordings (compared to 63.9% for the directional flute recordings here).

The differences between the cello and the flute recordings suggest that the influence of the room sound is somewhat dependent on the instrument and the musical material. As expected, the overall results are lower in almost every case using the directional RIRs compared to the omnidirectional RIRs.

It is certain that the influence of the room sound depends on many factors including the precise microphone position, the instruments and their radiation patterns, and the style of the compositions performed. However, this small experiment is useful to provide some context to the performance of these features in section 6.4 in predicting which room the musical performances took place in.

6.6 Discussion

The goal of this chapter was to take a more experimental approach towards music performance analysis informed by musicological principles. In contrast to the more generalized approach in chapter 5, this method sought to identify specific dimensions which are considered to be important in executing a historically appropriate baroque performance. These dimensions include phrasing (which is accomplished through manipulations of tempo and loudness), tone production (which is strongly correlated with timbre), and vibrato (a manipulation of pitch during a note's duration). While the effectiveness of the designed features to capture these exact musical properties is not precisely known, they were able to meaningfully differentiate between different styles of baroque interpretation which were previously confirmed by both musicologists and listeners to vary among these dimensions (Fabian and Schubert, 2009).

When applied to recordings from the main experiment, these features revealed that, among phrasing and tone production, there were some notable differences in the performances between the two rooms. These differences were more significant among the violists and flutists compared to the theorbists. This aligns with the analysis in chapter 5 which found similar ambiguities among the theorbists' performances and was likely due to the fact that the theorbists had difficulty performing the selected compositions without making mistakes. No significant difference was found in the usage of vibrato as a function of the room.

These results are a promising first step and help suggest areas of interest for further analysis. However, the potential for some of these features to be influenced by factors that are not of interest, such as the sound of the room's acoustics present in the signal, was significant. This influence may give the false impression that performances varied as a function of the room more than they actually did.

In order to better understand the extent to which various features may have been influenced by this factor, a small experiment was performed which compared anechoic recordings convolved with different RIRs meant to imitate the source-receiver configurations in the two rooms used in the main experiment. This test found that the classifiers were influenced by the room sound present in the recordings, however, this influence was, in most cases, rather modest and the results in section 6.4 all outperformed the lower bound suggested by this experiment.

While this method of performance analysis was able to reduce the question to only a few parameters, the problem of interpreting these resulting components remains. For example, the "phrasing" feature is really just a description of the variation of tempo and loudness within musically meaningful durations, and its correlation with any specific dimension of phrasing

(such as “legato” versus “detached”, for example) is unknown.

One aspect this analysis approach did not explore was how all of these features perform when combined. Most of these features were classified at an observation level. On a larger dataset, these could be combined with global descriptors to classify an entire performance. This could be useful for characterizing a larger set of baroque performances for use in cluster analysis or recommendation algorithms, for example.

One issue with developing this analysis framework on a small set of recordings of Bach violin pieces is the obvious risk of over-fitting to an unrepresentative dataset. Conclusions which appear to be well-founded when drawn from these results may not hold up when applied to broader contexts. Ultimately, without larger datasets of baroque music annotated along specific expressive dimensions, it is difficult to create a truly robust and dependable analysis framework. Still, the approach outlined in this chapter suggests that there are some benefits to a tailored approach of music performance analysis over a more generalized approach.

Chapter 7

Listening test

In the previous two chapters work was done to systematically identify differences in performance style among recordings of baroque performances. While these measures were able to indicate how different one set of performances was from another among specific dimensions, without a ground truth, the success of these methods is difficult to assess. Determining a definitive ground truth for something as complex and aesthetically charged as performance style is nearly impossible, but a listening test could provide some useful subjective data against which to compare the already obtained objective data. For example, consensus among listeners about certain musical or aesthetic judgments can help contextualize and potentially strengthen findings from the objective analyses from the previous chapters.

A listening test was designed to obtain perceptual judgments in the categories relevant to the features extracted in chapter 6 (namely, phrasing, tone production, and vibrato) in order to more directly elucidate connections between the objective measures extracted in that chapter and listeners' perception. These categories have already been deemed as essential to communicating baroque historically informed performance (HIP), which is an important factor in considering the overall success of these performances (Fabian and Schubert, 2009). The results from the listening test can provide further measure on the success of these features to capture what they were designed to capture. Furthermore, when viewed within the two acoustic contexts, these results can also offer meaningful insight into how the performance

Table 7.1 Performances selected for inclusion in the listening test (indicated as [musician number]-[take number]).

Flute					
		Extreme		Average	
		SdN	Amphi	SdN	Amphi
Phrasing	Animé	3-3	1-2	2-3	2-1
	Gravement	1-2	1-1	3-2	2-2
Tone	Animé	2-1	3-1	3-1	3-2
Production	Gravement	2-3	1-1	3-1	2-1

Viol					
		Extreme		Average	
		SdN	Amphi	SdN	Amphi
Phrasing	Couplet 12	4-2	4-3	3-2	3-2
	Couplet 18	2-2	2-3	1-2	3-3
Tone	Couplet 12	2-3	4-1	1-3	1-2
Production	Couplet 18	1-1	4-3	3-2	2-3
Vibrato	Couplet 12	2-3	4-2	3-3	1-2
	Couplet 13	1-1	1-2	4-2	4-1

style was perceived as differing as a function of the room among these musically meaningful dimensions.

7.1 Methodology

This section details the methodology for the selection of audio examples, as well as the design of the listening test, including the motivation behind the terminology used. The methodology was strongly influenced by the listening test performed in Fabian and Schubert (2009), which formed the foundation for the previously detailed verification of the baroque analysis framework (see section 6.3).

7.1.1 Recording selection

Nearly 200 recordings were made between the two real rooms during the main experiment (10 musicians, two rooms, and three repetitions of three compositions). A listening test including

all of these recordings would be too burdensome for participants, so a subset of recordings was chosen.

Because of the generally subtle differences in performance style between the two rooms, a random selection of these recordings would have likely resulted in a generally small effect size. As such, rather than selecting performances at random, the selections were curated to include two types of performances: one which represented an average performance in that room for that composition and feature type (from here on referred to as "average" performances), and one which represented the performance in that room which was most different from the performances in the other room for that composition and feature type (here referred to as "extreme" performances). The intention was to measure the maximum conceivable effect of the room by restricting analysis to the extreme performances relative to more representative effects via average performances.

The data was partitioned into smaller subsets of individual musicians and compositions. Within each subset and feature set, the Mahalanobis distance[1] was calculated between the cluster centroid of one performance's group of features (either phrasing, tone production, or vibrato) and the cluster of another performance's group of the same features. All performances within each subset were pairwise compared using this distance measure. The data for each composition was treated separately; no inter-composition distances were calculated. Performances that represented the largest inter-room distance were chosen as the extreme examples while performances that represented the smallest intra-room distance were chosen as the average examples.

The selected performances are listed in table 7.1. The theorbo recordings were excluded, since it was previously found that mistakes were common throughout these performances, rendering these recordings unreliable for such analysis. Only two of three compositions were chosen for the flute and viol, so as to reduce the number of total examples. These compositions were chosen based on their capacity to showcase the desired features. The selection process resulted in some examples being selected twice, resulting in 15 examples for the flute and 21 examples for the viol. Couplet 13 was chosen as the second piece for the vibrato features as the presence of vibrato in Couplet 18 was so low. The duration of each recording ranged from approximately 30 s to 60 s. All recordings within each instrument group were normalized to have the same root mean square (RMS) level.

[1]First introduced by Mahalanobis (1936), this is a distance measure commonly used in multivariate data that takes into account the correlations within the data.

Table 7.2 Listening test rating categories in both english and french. The right side descriptions in each category are associated with a more historically-informed baroque playing style.

Phrasing		Phrasé / articulation	
Continuous	Articulated	Continu	Articulé
Strict	Flexible	Strict	Flexible
Mechanical	Varied	Mécanique	Varié
Tone production		**Qualité et production du son**	
Forced/intense	Light	Appuyé/intense	Léger
Muddy	Clear	Embrouillé	Clair
Straight	Uneven	Stable	Inégal
A lot of vibrato	No vibrato	Beaucoup de vibrato	Sans vibrato
Baroque expressiveness		**Expressivité baroque**	
Not at all	Very	Pas du tout	Très

7.1.2 Design

Three significant musical dimensions for communicating baroque expressiveness from Fabian and Schubert (2009) and examined in chapter 6 are: phrasing, tone production, and vibrato. Stylistic tendencies within each of these dimensions which are typical of baroque HIP have already been discussed in section 2.1.3, but will be reiterated briefly here. Within phrasing, a historically-appropriate baroque performance is typically broken into smaller detached and articulated phrases of well defined metric groups with clear separation. An appropriate tone production within baroque HIP should be relatively light and transparent, with a shallow and limited use of vibrato and incisive articulation.

It is important to note that the characteristics of historical baroque performance discussed in this analysis are simplified for clarity. For example, while it remains broadly true that phrasing in historically informed performances of baroque music tends to be articulated rather than continuous, it is still possible for continuous phrasing to be valid in certain contexts of historical baroque performance. The same holds true for all of the considered parameters. Furthermore, the implementation and perception of expressive devices in music can be strongly influenced by the composition and instrument (Fabian, Schubert, and Pulley, 2010), and this

interaction has not been very well studied.

The test was administered through an application designed using MATLAB App Designer. The test was divided into three sections; a preliminary training section in which the participant was given two examples to judge in order to familiarize themselves with the interface, and one section for each instrument. These last two sections were presented in a randomized order, and, within each section, the order of the audio examples was also randomized. Two audio examples were duplicated for both the flute and viol in order to have a measure of the repeatability of the participants' ratings. The participants were free to take a short break in between the flute and viol sections if desired. The average duration of the test was between 45 and 60 minutes.

Participants were asked their age, level of general education (currently pursuing bachelor's degree, currently pursuing master's degree, or completed master's degree), and level of familiarity with baroque HIP (not very familiar, familiar, very familiar).

The main interface has a play/pause and stop button for the current audio example and lists eight categories on a 7-point (from -3 to +3) semantic differential scale for the users to rate. The rating scales were presented as continuous sliders. The participants were first given two training examples whose responses were not included in the analysis in order to become familiarized with the interface. The participants were allowed to revisit prior examples and modify their answers if desired.

The terms used are listed in both english and french in table 7.2. The three broad categories are phrasing, tone production (which includes vibrato), and baroque expressiveness. Vibrato was only made available for the viol examples. These terms were decided upon after a review of several papers examining listener perception of baroque performance (Fabian and Schubert, 2009; Schubert and Fabian, 2014, 2006). In general, the right side column of descriptors correlate with performance characteristics which tend to be more baroque appropriate. A definition of baroque expressiveness was provided during the training portion of the test: *a performance which adopts stylistic attributes which are characteristic of historically-informed baroque performance practice.* These terms were translated into french with the assistance of a french musicologist.

7.1.3 Participants

Twenty participants took part in the test. All participants were required to have some formal musical training at the university level and some familiarity with baroque HIP. Most participants were recruited from music students at the Clignancourt campus of the Sorbonne University. The experiment protocol was approved by the Research Ethics Committee (*Comité*

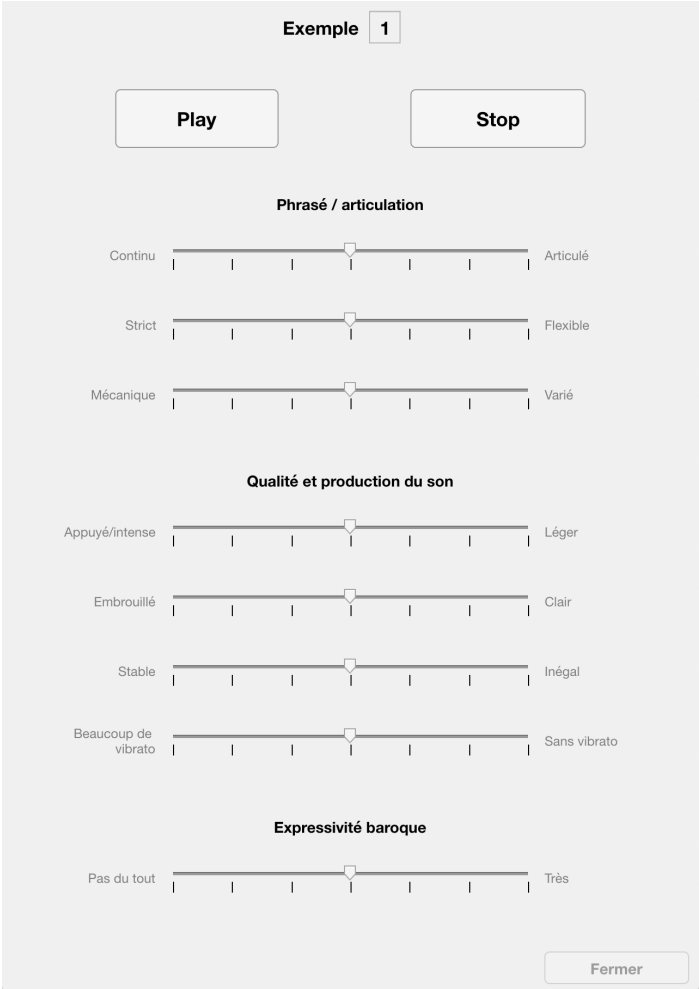


Figure 7.1 A screen capture of the listening test interface.

d'Éthique de la Recherche) at Sorbonne University (CER-2022-ELEY-EVAATest).

The average age of the participants was 24.8 (SD: 6.7). Eight participants reported some familiarity with baroque HIP, 11 participants reported familiarity with the subject, and one participant reported they were very familiar with it. Nine participants were bachelor's students, six participants were master's students, and five participants had obtained a master's degree.

7.2 Results

Figures 7.2 and 7.3 show box plots of performances in each category, grouped by room. Because the data generally did not follow normal distributions, and because the sample sizes were not always the same between the two rooms, a nonparametric Mann-Whitney U test was used to identify whether there were significant differences in the ratings of the examples from the two rooms. The Mann-Whitney U test treats the data as two independent samples, which was preferable to a repeated measures test as there was not a direct relationship between the recordings chosen in the two different rooms. This test was applied to several data subsets: the "average" examples, the "extreme" examples, and all examples together. The null hypothesis for the Mann-Whitney U test is that there is no difference between listener ratings based on the room. The resulting p-values are reported in tables 7.3 and 7.4.

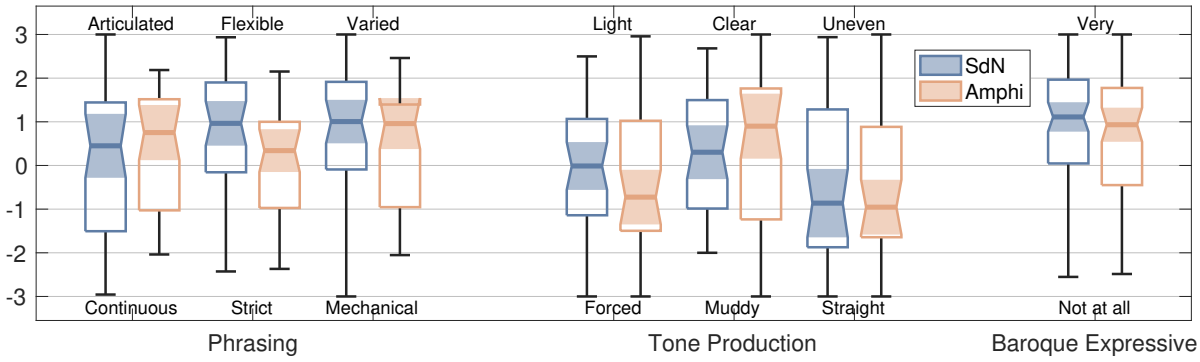
A Cliff's delta statistic accompanies p-values at the 5% significance level or greater. Cliff's delta takes on the range of $-1 \leq \delta \leq 1$ which indicates the amount of overlap between two samples. A resulting positive δ value would indicate the performances were perceived as more baroque appropriate in that dimension in the Salon des Nobles, the baroque-era room.

As previously mentioned, two duplicates were included in the examples for both instruments in order to examine the reproducibility of the participant responses. The average absolute difference between the ratings across all categories of these duplicates for the flute was 1.11 (on the seven-point scale) with a standard deviation of 0.54. For the viol, the average absolute difference was 1.06 with a standard deviation of 0.41.

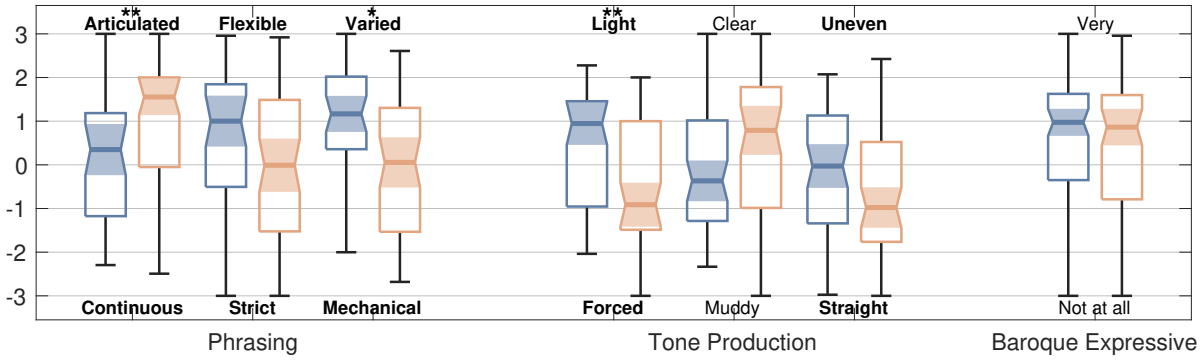
7.2.1 Flute results

No significant differences were observed among the average flute performances according to the Mann-Whitney U tests (see table 7.3). However, among the extreme performances and among the combined set of performances, significant differences were found in every category except for the muddy—clear dimension of tone production and the baroque expressiveness rating.

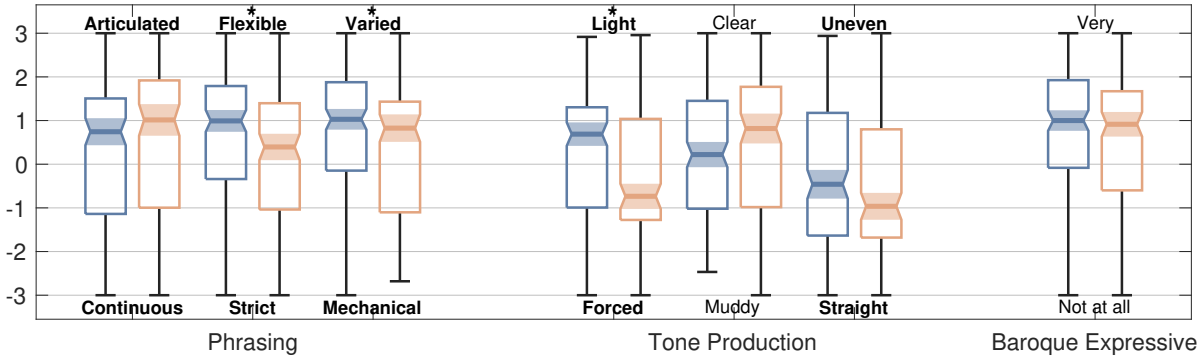
Within the continuous—articulated dimension of phrasing, the performances in the Salon des Nobles were rated as significantly more continuous, while those in the amphitheater were rated as more articulated ($p = .029$, $\delta = -0.14$, for all examples). This is contrary to claims by some of the flutists in their questionnaire responses (see section 5.4.2) that they intentionally tried to extend the duration of notes when playing in the amphitheater to compensate for the lack of acoustical decay in that hall, indicating that the flutists were perhaps not able to achieve their intended performance goals.



(a) Average examples.



(b) Extreme examples.



(c) All examples.

Figure 7.2 Box plots of listening test results for flute performances showing (a) average examples, (b) extreme examples, and (c) all examples. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima. Bold labels, asterisks (*) and double asterisks (**) indicate p-values of $\leq .05$, $\leq .01$, and $\leq .001$, respectively, according to a Mann-Whitney U test.

Table 7.3 Listening test results for flute examples showing p-values (and Cliff's δ where significant p-values were found) of Mann-Whitney U tests of responses for performances in the two different rooms. Bold, single asterisk (*****) and double asterisk (******) indicate significance at the .05, .01, and .001 levels, respectively.

	Average		Extreme		All	
Phrasing	p	δ	p	δ	p	δ
Continuous—Articulated	.729	-	< **.001****	-0.39	**.029**	-0.14
Strict—Flexible	.086	-	**.050**	0.23	**.002***	0.20
Mechanical—Varied	.408	-	**.006***	0.34	**.002***	0.20
Tone Production						
Forced—Light	.341	-	< **.001****	0.34	**.007***	0.17
Muddy—Clear	.893	-	.120	-	.216	-
Straight—Uneven	.564	-	**.021**	0.25	**.022**	0.14
Baroque expressive	.282	-	.295	-	.152	-

An area where there was agreement between the questionnaire responses and the listener ratings is the forced/intense—light dimension of tone production. Flutists described the need to "project further" or "dig deeper into the dynamics" in the amphitheater, and one would expect this to result in a more "forced/intense" tone. This is exactly what was indicated by the participants in this listening test, with a high significance and moderate effect size ($p = .007$, $\delta = 0.17$, for all examples). This indicates that the intention to project more in the amphitheater was perceived by listeners. Furthermore, a light tone is associated with a historical baroque playing style meaning that, at least in this dimension, the flutists were perceived as producing a more baroque-appropriate tone in the Salon des Nobles than in the amphitheater.

Highly significant differences and moderate effect sizes were found within the other two dimensions of phrasing; strict—flexible and mechanical—varied ($p = .002$, $\delta = 0.20$, for all examples in both dimensions). Listeners found the performances in the Salon des Nobles to be more flexible and varied, while those in the amphitheater were judged to be more strict and mechanical. Flexible and varied phrasing is associated more with a historical baroque performing style, indicating that, among these dimensions, the flutists' performances were

generally perceived as being more baroque appropriate in the Salon des Nobles.

There appears to have been no difference in perceived baroque expressiveness as a function of the room within the flute performances, regardless of which subset of performances is examined.

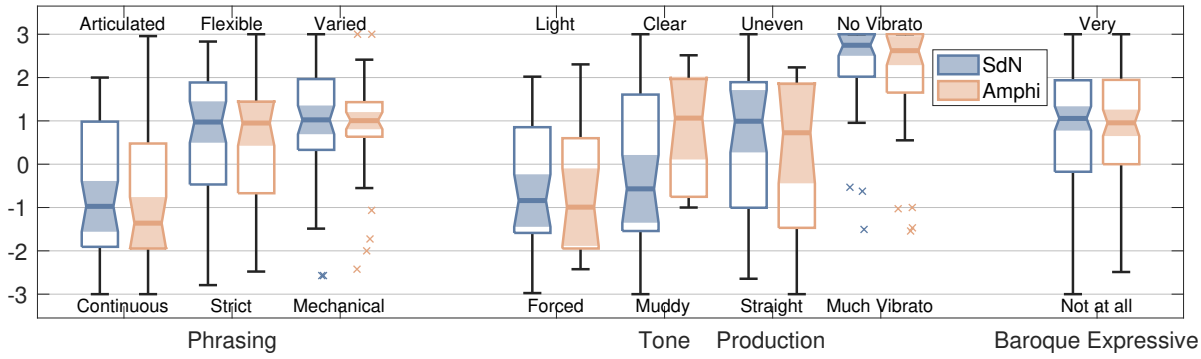
7.2.2 Viol results

As with the flute examples, no significant differences were observed among the average examples of the viol performances (see table 7.4). Among the extreme examples, significant differences were found within the mechanical—varied and strict—flexible dimensions of phrasing and within the straight—uneven dimension of tone production. The results using all examples showed significant differences only in the three phrasing dimensions.

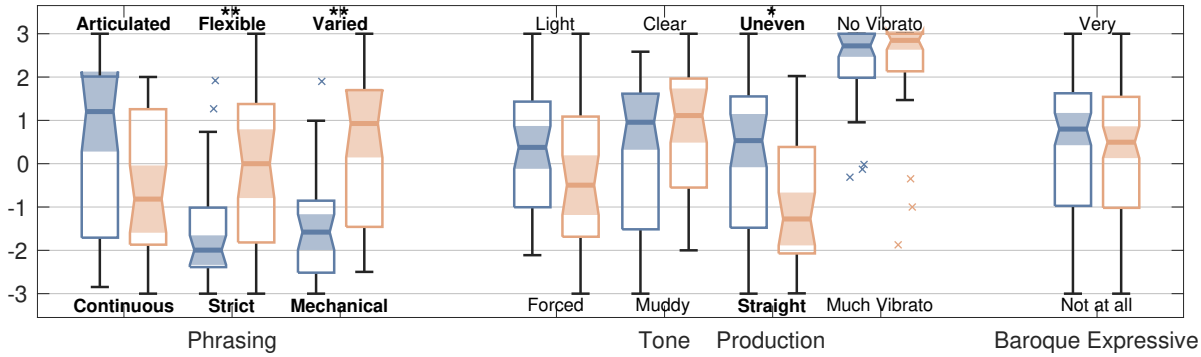
Among the extreme examples, the viol performances were judged to be significantly more strict ($p < .001$, $\delta = -0.44$) and mechanical ($p < .001$, $\delta = -0.56$) in the Salon des Nobles and more flexible and varied in the amphitheater.

However, when looking at the results for all examples, the effect sizes are in the opposite direction for these two dimensions of phrasing. That is, the performances in the Salon des Nobles were rated as exhibiting slightly more flexible ($p = .003$, $\delta = 0.16$) and varied ($p = .005$, $\delta = 0.15$) phrasing, albeit with much smaller effect sizes. This indicates that, among these dimensions and all examples, these performances were perceived as being slightly more baroque appropriate in the Salon des Nobles. Within the remaining dimension in the phrasing category, continuous—articulated, the performances in the Salon des Nobles were rated as more continuous (and therefore *less* baroque appropriate in this dimension) than those in the amphitheater ($p < .001$, $\delta = -0.20$ for all performances).

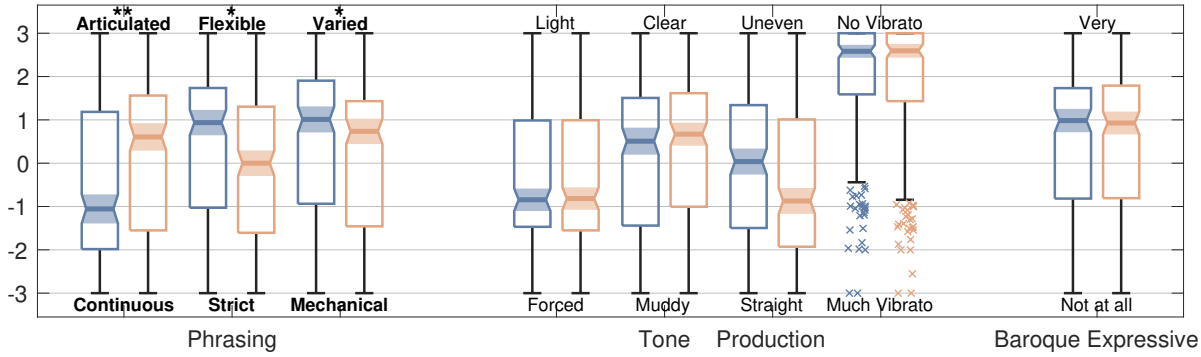
The only significant difference observed within the tone production category was among the straight—uneven dimension with the extreme examples ($p = .005$, $\delta = 0.33$). These ratings indicated that performances in the Salon des Nobles were rated as having more uneven tone production. When including all examples, this trend was still observed but was not statistically significant ($p = .066$).



(a) Average examples.



(b) Extreme examples.



(c) All examples.

Figure 7.3 Box plots of listening test results for viol performances showing (a) average examples, (b) extreme examples, and (c) all examples. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima. Bold labels, asterisks (*) and double asterisks (**) indicate p-values of $\leq .05$, $\leq .01$, and $\leq .001$, respectively, according to a Mann-Whitney U test.

Table 7.4 Listening test results for viol examples showing p-values (and Cliff's δ where significant p-values were found) of Mann-Whitney U tests of responses for performances in the two different rooms. Bold, single asterisk (*****) and double asterisk (******) indicate significance at the .05, .01, and .001 levels, respectively.

	Average		Extreme		All	
Phrasing	p	δ	p	δ	p	δ
Continuous—Articulated	.254	-	**.050**	0.26	< **.001****	-0.20
Strict—Flexible	.658	-	< **.001****	-0.44	**.003***	0.16
Mechanical—Varied	.625	-	< **.001****	-0.56	**.005***	0.15
Tone Production						
Forced—Light	.520	-	.128	-	.692	-
Muddy—Clear	.071	-	.075	-	.131	-
Straight—Uneven	.312	-	**.005***	0.33	.066	-
Vibrato	.562	-	.519	-	.878	-
Baroque expressive	.841	-	.966	-	.614	-

As expected, based on the vibrato findings in section 6.4.3, there was little to no perceived difference in vibrato usage among the performances in the two different rooms. The vibrato ratings indicated that there was very little perceived vibrato overall, aligning with the previously reported objective measures.

There appears to have been no difference in perceived baroque expressiveness as a function of the room within the viol performances, regardless of which subset of performances is examined.

7.2.3 General discussion

For both instruments, there were no significant differences observed when restricting the analysis to the "average" performances. These average performances were selected based on their proximity in a multivariate feature space to all other performances in the same room, making them the most representative performances in that room, for a specific feature and composition. The lack of significant differences found as a function of the room for these performances suggest that average performance adjustments are quite difficult for a listener to discern.

When restricting the analysis to the "extreme" performances only, or to all performances,

significant differences were found for both instruments in several dimensions. The extreme performances were selected to be the most different from the average performance in the other room, for a specific feature and composition. This suggests that listeners are able to perceive performance changes on this scale.

Based on the questionnaire responses (see section 5.4.2), one would have expected the flutists' performances to be perceived as having more continuous phrasing and a more forced/intense tone in the amphitheater. While listeners did perceive a more forced tone in the amphitheater, they rated the phrasing there as more articulated. This suggests that the musicians were not successful in communicating all of their performance intentions to the listeners. The two primary performance intentions of the flutists in the amphitheater, according to questionnaire responses, were to increase their projection and to lengthen the notes to their full duration. However, an increased projection requires more effort and breath support, and would therefore likely render it more difficult to sustain notes to their maximum duration, meaning these two efforts were somewhat opposed to each other. Therefore, it is probably that only one of these intentions was achieved and communicated to the listener.

The flutists' attempt to increase projection in the amphitheater may have resulted in their performances being perceived as more strict and mechanical in terms of phrasing. This focus on projection may have made it harder for them to focus on other aesthetic concerns, such as phrasing, which could have led to a more mechanical interpretation.

Significant differences were found in the phrasing category for both the extreme examples and all examples for the viol, but the effect sizes for these two data subsets were in opposite directions. This suggests that the extreme examples selected to represent the phrasing features for the viol were perceived very differently from the other performances. This discrepancy may be partly due to a strikingly different interpretation of Couplet 12 by violist 4 (see section 5.4.1.4) who represented the extreme phrasing examples for both rooms for this composition. This musician played the piece, which consisted only of chords, as single, abrupt strokes while all of the other musicians arpeggiated them. This violist performed the entire piece as single strokes in the Salon des Nobles, whereas in the amphitheater they arpeggiated the second half of the piece. This aligns with the results that show the extreme performances in the Salon des Nobles as exhibiting significantly more strict and mechanical phrasing. The recording selection process assumed that all interpretations were reasonable, and therefore the extreme examples would represent the maximum reasonable performance changes exhibited by musicians. However, the performances by this violist may have exceeded expected performance changes. Figure 7.4 illustrates how distinct this violist's performances were, compared to those

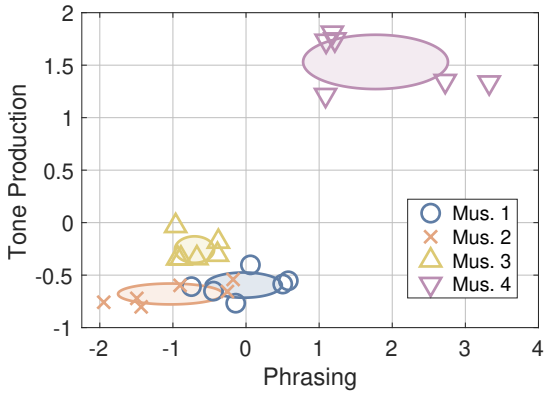
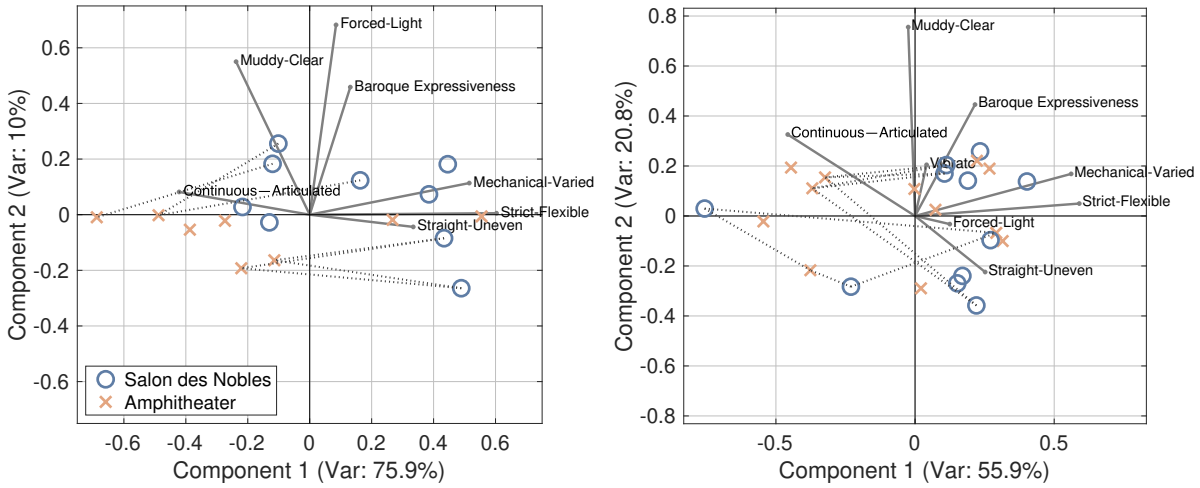


Figure 7.4 The distribution of objective data of violist performances of Couplet 12 in both rooms. The x-axis is the top principal component of the phrasing features while the y-axis is the top principal component of the tone production features. Individual data points as well as the standard deviation around their means are shown. The distributions show how different the interpretation of violist 4 was from the other musicians.

of the other musicians, as indicated by objective features. It is notable that this violist reported having 3 years of professional experience, whereas the remaining three violists had self-reported 31-41 years of experience (mean: 37.3). It is possible that this relative difference in experience may have been partly responsible for this unconventional interpretation. While this performance stood out as being audibly different than the others, it was still included in favor of an automated approach which would be free of selection bias.

The lack of significant difference observed in baroque expressiveness between the two rooms does not necessarily indicate that listeners were not able to discriminate within this global parameter. As shown in Fabian and Schubert (2009), listeners were able to discriminate along this broad dimension, however, in that study, the musical examples were chosen to represent a wide range of baroque playing styles and therefore a wide range of baroque expressiveness. In this study, all of the performances generally adopted the same baroque HIP style, so any differences within the baroque expressive parameter would be expected to be rather small.

Due to the fact that significant differences were found in many of the more narrowly-defined parameters within phrasing and tone production, most of which were in the direction of being more baroque expressive in the Salon des Nobles, one might expect a similar difference to be found in the baroque-expressive parameter. This was not the case, however, suggesting that it is easier for listeners to provide consensus among more narrowly-defined parameters than on



(a) Flute **(b)** Viol

Figure 7.5 Biplots of the first two principal components resulting from a principal component analysis (PCA) of the listening test responses of the (a) flute and (b) viol examples. Dotted lines show connections between extreme performances of the same compositions in different rooms (excluding vibrato).

global parameters such as baroque expressiveness.

7.2.4 Principal component analysis of ratings

The results of the listening test were subject to principal component analysis (PCA) to explore what makes up the most salient perceptual dimensions. Figure 7.5 shows biplots of the resulting first two components of this analysis for the flute and viol results, analyzed separately. Dotted lines are included to show connections between extreme examples of phrasing and tone production features of each composition between the two rooms. Connections among the extreme vibrato examples in the viol were left out to simplify the figure, since no difference was found within this dimension as a function of room. The first two components combined are responsible for 86.9% and 76.6% of the variance for the flute and viol examples, respectively.

The flute results (fig. 7.5a) show fairly clear separation between the rooms among the first component, which is mostly made up of the phrasing parameters, along with the straight—uneven dimension of tone production. The second dimension appears to consist of factors more related to tone production, including the forced—light and muddy—clear dimensions. The baroque-expressive dimension is made up of somewhat of both dimensions, but is slightly more correlated

with the second component. It is notable that the location of the baroque-expressive dimension suggests that a performance is perceived as more baroque-expressive when it is judged to exhibit a lighter tone and more varied/flexible phrasing. This aligns well with musicological expectations.

The separation of classes (rooms) appears to be more distinct among the first component which is responsible for much more of the overall variance (75.9%) than the second component (10.0%). The third component was responsible for only 7.0% of the variance. Most of the variation of the extreme performances seems to be along the first component, as indicated by the dotted lines. This suggests that the first component is the primary dimension along which performances varied as a function of the room.

The viol results (fig. 7.5b) show that the strict—flexible and mechanical—varied dimensions of phrasing are strongly correlated with the first component while the muddy—clear dimension of tone production is most strongly correlated with the second component. The baroque-expressive dimension seems to be made up of both components roughly equally. These results suggest that the viol performances were perceived as being more baroque-expressive when the performances were judged to have a clearer tone and more varied/flexible phrasing. This is also compatible with musicological expectations. There does not appear to be a very clear separation of rooms among either of these components as there was within the flute results. The top two components contribute fairly significant variance each (55.9% and 20.8%), while the third component contributes only 12.2% of the overall variance. The variation of the extreme performances tends to be mixed along the two components, as indicated by the dotted lines. This suggests that there is no primary dimension along which the viol performances varied as a function of the room.

For both the flute and the viol, the strict—flexible and mechanical—varied dimensions appear to be the most active, being strongly correlated with the first principal component. This correlates with most significant differences being found in these dimensions. Furthermore, for both instruments, the mechanical—varied and strict—flexible dimensions of phrasing appear to be roughly opposed to the remaining phrasing dimension, continuous—articulated.

7.2.5 Comparison with objective features

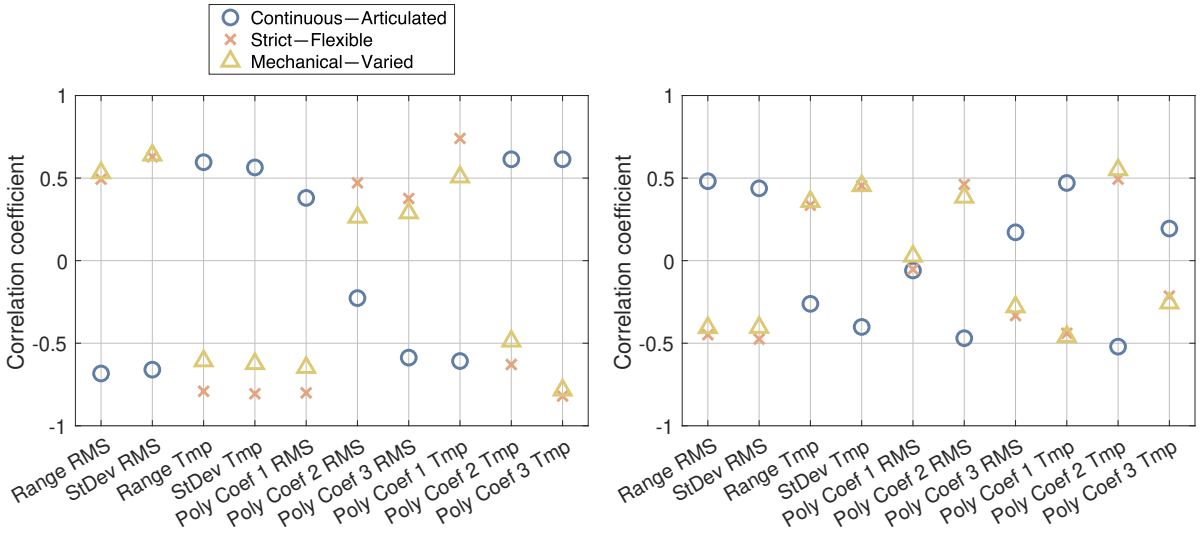
The subjective responses from the listening test were compared with the objective measures from the general analysis framework described in chapter 5 and the baroque analysis framework described in chapter 6. First, a subset of the objective features and subjective ratings were chosen (i.e. the phrasing features from the baroque analysis framework and the ratings of the

Table 7.5 Pearson correlation coefficients (r) between the sum of pairwise distances for all flute and viol performances of listener rating data (columns) and objective performance data (rows) from the baroque analysis framework (see chapter 6). Bold labels, asterisks (*) and double asterisks (**) indicate a p-values of $\leq .05$, $\leq .01$, and $\leq .001$, respectively. (B.E. = baroque-expressive.)

	Flute				
	Phrasing	Tone	B.E.	P.C. 1	P.C. 2
Phrasing	**0.60**	0.05	-0.04	**0.55**	-0.27
Tone production	-0.02	**0.78****	0.06	0.12	**0.56**
	Viol				
Phrasing	0.18	**0.42**	0.23	0.28	0.19
Tone production	**0.54***	0.34	0.21	**0.43**	0.09

phrasing dimensions from the listening test). Then, the median across all subjects for each piece and parameter was taken of the listening test ratings resulting in an $m \times n$ matrix where m is the number of performances and n is the number of rating categories. The mean of all objective features was taken for each piece resulting in an $m \times p$ matrix where p is the number of objective features. For both of these matrices and for each performance, m, the pairwise Euclidean distances to all other performances were calculated then summed, resulting in an $m \times 1$ vector for both the subjective responses and the objective data. Each value in this vector represents the distance of that performance to all other performances for that specific metric (either subjective ratings or objective performance data). Lastly, these two vectors were used to calculate a Pearson correlation coefficient (r). The coefficient r serves as a measure of similarity for how a specific set of objective measures and subjective ratings differentiate between performances. Results from these comparisons are reported in tables tables 7.5 and 7.6. The tone production features used in this analysis were those that were calculated using the original, unprocessed audio (as opposed to the harmonic or percussive components).

In addition to comparisons with the subjective rating categories of phrasing, tone production, and baroque expressiveness, the top two principal components from the PCA performed in section 7.2.4 were also included. These components were included because the most salient perceptual dimensions revealed by the PCA did not perfectly align with the two broad categories of phrasing and tone production. This offers another way to see how the objective data aligns with the perceptual ratings of the listeners. Because very little difference in vibrato was



(a) Flute **(b)** Viol

Figure 7.6 Correlation coefficients between subjective ratings in individual dimensions within the phrasing category and individual phrasing features from the baroque analysis framework of flute and viol performances. Tmp refers to the note-level tempo and RMS refers to the frame-wise RMS. The objective phrasing features are described in section 6.3.2.

found between the two rooms, according to both the objective measures and the subjective ratings, this parameter was left out of the following analysis.

The flute results in table 7.5 show strong correlations between the objective measures and the listeners' perception of the musical parameters they were intended to capture. The phrasing features show strong, significant correlations with the corresponding phrasing ratings while the tone production features show the same with their corresponding ratings. This indicates that these custom objective features are able to identify and isolate the perceptual dimensions they were designed to capture, at least for flute recordings. There is also some correlation between the phrasing features and the first principal component as well as the tone production features and the second principal component, adding further support that these are the most salient perceptual dimensions revealed by the questions posed. There are no significant correlations between the objective measures and the baroque-expressive ratings.

The results for the viol performances in table 7.5 show that there is a significant correlation between the phrasing features and tone production ratings, and also the tone production

features and phrasing ratings. A significant correlation was also found between the first principal component and the tone production ratings. One violist's unconventional interpretation (previously discussed in section 7.2.3) may have contributed to the unclear results in the viol performances. However, when these performances were removed, no significant correlations were found between objective measures and subjective ratings. This suggests that there is a problem with either listener consensus in evaluating viol performances or the effectiveness of the features in measuring their intended expressive parameters, despite the influence of the unconventional performances. An analysis of the average standard deviation of participant responses for each category, across all examples did not show a significant difference between instruments, however. This indicates that listener ratings for the flute and the viol were similarly consistent and therefore, the most likely reason for these results is that the objective features were simply not very effective at capturing their intended expressive performance parameters for viol recordings.

A direct comparison between the objective phrasing features and the subjective phrasing ratings for the flute and viol can be seen in fig. 7.6. The flute results (see fig. 7.6a) show that as the range and standard deviation of the intensity curves increase, the phrasing is perceived as more continuous, flexible, and varied. However, for the viol, the opposite trend is true (see fig. 7.6b). In both cases, the features derived from tempo curves seem to be negatively correlated with those derived from intensity curves. In general, the correlations observed in the flute examples tend to be stronger than those observed in the viol examples.

Using objective features from the general analysis framework proposed in chapter 5, correlations were found between the full feature set (reduced by PCA) and the phrasing and tone production responses as well as the first principal component from the flute performances (see table 7.6). Somewhat surprisingly, restricting the objective data to only tempo-related features resulted in fairly high correlations with the responses within the phrasing and baroque-expressive categories as well as the first principal component.

Observing the connection between these objective data and the listening test responses for the viol performances, a significant r was found only between the full objective feature set and the subjective phrasing and baroque expressive categories as well as first principal component (see table 7.6). These correlations show the power of the full feature set, however, unlike the baroque analysis framework, this feature set is not capable of isolating subjective parameters. Unlike the flute, no significant correlations were observed when restricting the objective data to fewer features.

Table 7.6 Pearson correlation coefficients (r) between the sum of pairwise distances for all flute and viol performances of listener rating data (columns) and objective performance data (rows) from the general analysis framework (see chapter 5). Bold labels, asterisks (*) and double asterisks (**) indicate a p-values of $\leq .05$, $\leq .01$, and $\leq .001$, respectively. (B.E. = baroque-expressive.)

	Flute				
	Phrasing	Tone	B.E.	P.C. 1	P.C. 2
All	**0.58**	**0.56**	0.11	**0.67***	-0.13
Loudness & tempo	0.47	-0.00	0.36	**0.50**	-0.34
Tempo only	**0.69***	-0.27	**0.54**	**0.62***	-0.28

	Viol				
All	**0.70****	0.06	**0.44**	**0.78****	-0.06
Loudness & tempo	0.05	0.11	0.21	0.21	-0.11
Tempo only	-0.19	0.08	0.15	-0.11	-0.17

7.3 Discussion

The purpose of this listening test was to gain insight into the perceived differences between performances in two separate rooms, and to determine if trained listeners' perceptions aligned with previously recorded objective differences. No significant differences were found when restricting the analysis to average performances. However, as previous studies have suggested that performance changes due to room acoustics are quite subtle, this outcome was not unexpected. When analyzing the extreme performances, or all performances together, a number of significant differences were found in several performance dimensions for both instruments as a function of room.

The most significant differences with the largest effect sizes were found within the phrasing category for both instruments. Performances in the Salon des Nobles were rated as being more baroque appropriate in the strict—flexible and mechanical—varied dimensions. However, within the continuous—articulated dimension, the performances were rated as being less baroque appropriate in the Salon des Nobles.

Within the flute results, some findings were consistent with performance changes reported by musicians, such as the tone production being judged as more forced in the amphitheater. However, ratings in the continuous—articulated dimensions did not align with the intended

performance changes reported by flutists.

There was fairly good agreement between the objective performance data and the listener ratings for the flute. There was a significant correlation between the proposed phrasing features and the listener ratings in the phrasing dimensions, while the same was true for the tone production features and their corresponding ratings. Additionally, the features showed almost no correlation with listener ratings in other categories. This is strong evidence to support the efficacy of these features to capture the expressive musical qualities that they were intended to capture, at least for flute performances.

The relationship between features and listener ratings was not as clear for viol performances. Tone production features correlated with ratings in the phrasing category and phrasing features correlated with ratings in the tone production category. These correlations, though significant, were not as strong as the correlations observed within the flute examples. These somewhat surprising results may be partially explained by an atypical interpretation of one of the violists, since, when these performances were removed, these correlations disappeared. Further analysis indicated that there was a similar consistency in responses for the two different instruments suggesting that the objective analysis was not very effective in capturing the expressive performance parameters of the violists.

This test provided evidence that there are some significantly different perceptual changes in performance style as a function of the room. Furthermore, listeners were able to perceive performance characteristics that were either reported by the performer, observed in objective parameters, or both. The most conclusive findings came from the flute examples, rather than the viol, suggesting that the previously used baroque analysis framework does not apply equally well to all instruments. Future research could focus on improving the objective analysis to be more robust to different instruments. The study also provided further insight into how listeners perceive baroque expressiveness, finding that it tends to be correlated with certain dimensions of phrasing and tone production. However, the specific dimensions may be somewhat dependent on the instrument. Additionally, the study found some evidence to support that the acoustics of the baroque-era room facilitated the performance of historical baroque music. Responses to the performer questionnaire suggested satisfaction with the acoustics of the Salon des Nobles, and listener ratings indicated that certain performance characteristics were rated as significantly more baroque appropriate in this room.

Chapter 8

Conclusion

The primary goal of this research was to investigate the role of room acoustics within historically informed baroque performance practice. However, the experiment protocol was designed not just to investigate this question, but also to study the impact of novel elements in an innovative auralization system. Importantly, the research allowed for the in-depth exploration of new methods of music performance analysis, which is an essential component of the larger research question.

While other studies have considered the effect of room acoustics on music performance in more general contexts, a central goal of this thesis was to concentrate on the less-studied domain of historical baroque music. An important justification for this was to provide some observational examples through which to contextualize existing research withinin historically informed performance (HIP) practice, where the performance space has generally not been a very important consideration. More broadly, this research can be valuable for teaching and performance efforts. Empirical data can help improve teaching strategies for dealing with different acoustics, and concert planners can be more knowledgeable of the acoustic needs of musicians, allowing them to make informed decisions regarding the settings for specific performance contexts. Furthermore, this effort provided useful data within the context of the nascent field of heritage acoustics which is important to support future efforts in sustaining and preserving intangible heritage.

While the focus on baroque HIP over mainstream music performance styles was an important aspect which set this research apart from previous efforts, there were also some disadvantages to the approach. For one, the findings from the study are not as widely applicable as a study on more mainstream music performance would have been. The lack of existing research in the field also made it difficult to contextualize the findings. And while the novel challenge of analyzing baroque music provided some useful opportunities (see chapter 6), it also generated some obstacles as some existing tools for analyzing music performance are not well-suited for baroque music. The uncommon historical baroque instruments used also created challenges, for example, when implementing their directional characteristics in the virtual acoustic environment (VAE) as discussed in section 3.1.1. Despite these challenges, the novel approach of focusing on baroque HIP provided valuable insights and opened up new avenues for future research in the field.

Within the experimental virtual archaeological-acoustics (EVAA) auralization architecture, which was detailed in chapter 3, there were several novel components which were intended to be evaluated during the experiment. These were (i) an auralization based on calibrated geometrical acoustics (GA) models, (ii) dynamic implementation of instrument directivity, (iii) and immersive adaptive visual rendering. Each of these components were discussed along with their intended outcome of improving the flexibility or immersion of the virtual environment with the ultimate end goal of making the environment as realistic and unobtrusive as possible. However, due to the many problems encountered in the auralization system, a full evaluation of all of these components was not possible and remains as further work.

Finally, the resulting recordings offered some useful opportunities to develop and compare various approaches of music performance analysis, discussed in chapters 5 and 6. Broadly speaking, these two methods represented a generalized approach (chapter 5) and a customized framework developed specifically for historically informed baroque performance (chapter 6). A listening test was performed using the recordings from the main experiment, to better understand the perceptibility of changes in performance style as a function of the room as well as to shed light on the efficacy of the proposed analysis frameworks (see chapter 7).

8.1 Summary of findings

The conclusions from the two primary components of this research are summarized here. Section 8.1.1 discusses the outcomes from the VAE assessment while section 8.1.2 discusses the main findings regarding the effects of acoustics on the musicians' performance.

8.1.1 Virtual environment assessment

The EVAA auralization system developed for this project was not a sufficient replacement for the real acoustic environment. It received fairly low ratings from participants in various quality assessments and received significantly different scores in acoustic rating questionnaires than the corresponding real environments (see section 4.2). An assessment of the source of these faults found that the primary problem was that the VAE was producing an unnatural coloration described by many participants as "metallic."

The auralization method used in this study, which was based on calibrated GA models, had not been used before in a study investigating the experience of musicians in different acoustic settings. However, it was difficult to assess the success of this method due to the other problems in the auralization. There is some evidence that certain acoustic attributes, such as reverberance, were reproduced well, while others, such as timbre, were not (see fig. 4.1). The concept of basing auralizations on calibrated GA models offers clear advantages to previous methods, such as the ability to change the source-receiver configuration and the ability to modify the acoustics of the space under study, and therefore deserves further consideration. An experiment that reduces the auralization system to a setup which has already been done successfully (such as many of the studies discussed in section 2.3) could allow for better evaluation of the effectiveness of this method compared to an auralization based on measured room impulse responses (RIRs) or uncalibrated GA models such as those previously discussed in section 3.7.

The capability of the novel dynamic directivity component was challenging to evaluate due to limited movement by the musicians during their performances. In order to form a more accurate assessment of its effectiveness, it may be beneficial to choose a different use case in the future. Recent research has suggested that humans are not particularly adept at perceiving source directivity in reverberant environments in first person use cases (Frank and Brandner, 2019). This suggests that a simpler approach may be possible. However, more research is needed to fully understand the perceptual limits of source directivity in auralizations where the source and receiver are coincident, such as in first-person experiences or, in this case, where the musician is performing and listening simultaneously, since there currently are not many studies devoted to this question.

The visual models were impressive when viewing the offline-rendered versions, however, the limitations in the real-time rendering process were apparent, and the resolution offered by the projector, compared to the size of the image, could be improved to enhance the overall realism of the visual experience, though at a significant financial cost. This visual component was rated unfavorably by the participants and even considered by some as an unwelcome distraction (see

section 4.3). A quieter projector may improve future iterations of this setup since, despite attempts to mitigate its sound, the projector noise was a source of complaints. As discussed in section 3.6.1, there are advantages and disadvantages to including such a component in an auralization system, and while it can be an asset in certain contexts, the decision to include one must be made carefully and judiciously.

8.1.2 Effect of acoustics on performance

The findings from this study indicated that there was no overarching strategy for adapting playing style to acoustics that transcended musician, instrument, and composition. However, when looking at smaller subsets of the participants, some patterns were revealed. For example, all flutists had a fairly similar assessment of the acoustics in the amphitheater. They complained of a lack of acoustical support or return from the room, likely due to the hall's shorter reverberation time in the middle frequency range (see fig. 3.4). By contrast, their assessment of the acoustics of the Salon des Nobles was generally very positive. In response, the flutists claimed to develop strategies which included trying to extend the duration of the notes to their full extent to compensate for the lack of reverberance and by increasing their projection. Evidence from the listening test indicates that listeners were able to hear a difference in tone production resulting from the forceful projection necessitated by the acoustics in the amphitheater. However, there is also evidence from the listening test that contradicts the strategy of lengthening notes in the amphitheater. Generally, more evidence was found to support a change in tone production by the flutists in response to the amphitheater's lack of acoustical support, suggesting that only some of the flutists' intentions were communicated to listeners.

The critique of the amphitheater's lack of acoustical support was shared by two of the three theorbists, although their intended strategies in response to the acoustics differed. One theorbist claimed that they felt pressure to increase their tempo as a way to shorten the empty space between notes, although they attempted to resist this inclination by regulating their tempo more strictly. The other theorbist adopted a strategy similar to the flutists, by trying to let the strings ring more freely as an attempt to fill the space left empty by the lack of reverberance.

Three of the four violists also made some reference to the difference in resonance or reverberation time between the two rooms, although their stated adaptation strategies varied. One violist adopted the same strategy as some of the other musicians, attempting to elongate notes in response to the amphitheater's acoustics. The other two violists were not specific about how they adjusted their playing in response to the acoustic differences between the two rooms.

Most of the musicians described how they felt compelled to adjust their playing due to the acoustics in the amphitheater, whereas in the Salon des Nobles they generally felt supported by the acoustics and therefore could express themselves freely, allowing them to focus on producing an expressive performance. This forms a fairly strong case that the acoustics of the Salon des Nobles facilitated the performance in the style of baroque HIP while the acoustics of the modern amphitheater did not.

It is important to note that the most common adaptation strategies to the acoustics in the amphitheater—elongating the duration of the notes and playing with a more intense tone—run directly counter to what is generally accepted as an appropriate playing style in historical baroque performance practice, where it is commonly advised to play in a detached, rather than legato, playing style and with a lighter tone. As such, not only did the amphitheater necessitate an intentional adaptation of playing style due to its acoustics, but it also forced the musicians to adopt a less baroque-appropriate playing style, a fact which was partially confirmed by the listening test.

While the adaptation strategies described above are fairly consistent, finding direct correlates in the measured objective data was not straightforward. In some cases, there was evidence, such as indications from a Friedman test, that supported a significantly different playing style where the musician described a specific change. However, in several cases, some significant differences were found that had no correlation with the musician's described performance intentions. Ultimately, without explicit information from the musician regarding their performance intentions (which was not always available), it was difficult to say whether any measured changes were due to the acoustic differences between the rooms, or simply due to the natural variation in performance style from one day to the next.

Both strategies for analyzing music performance were useful. A broad analysis strategy, like what was used in chapter 5, was useful for identifying differences among different sets of performances, which helped establish strategies for further analysis. This in-depth analysis, however, still requires a great deal of human intervention to evaluate and contextualize the specific differences measured.

The baroque analysis framework, described in chapter 6, shows potential as a music performance analysis strategy. For example, this framework was capable of identifying differences at a higher level, utilizing more musically meaningful parameters such as vibrato, which is important to baroque performance practice, even if not much difference in vibrato was observed between the two rooms. The results from the listening test showed that this framework was able to capture and isolate the musical dimensions of phrasing and tone production, at least

for flute examples. In order to create a more robust and dependable analysis framework, more data would be needed in the form of recordings of baroque music with playing styles annotated and confirmed by informed listeners.

Lastly, the listening test provided useful information on the perceptibility of the changes in playing style reported by musicians and those indicated through objective measures. Some of the results supported previous findings, like the judgments of the flutists' tone production while some results contradicted previous findings, such as some ratings of the flutists' phrasing style. More generally, the listeners found more significant differences with regards to the phrasing style as a function of the room. These changes generally indicated that the musicians performed with a more baroque appropriate phrasing in the Salon des Nobles. This supported previous findings in section 6.4 that indicated there were larger differences in phrasing style than in tone production as a function of the room. No significant differences were found in terms of general baroque expressiveness as a function of the room, although the flute results did show that the performances in the Salon des Nobles were rated as being slightly more baroque-expressive, at a near significant level.

The listening test results also confirmed the efficacy of the baroque analysis framework to capture the features they were designed to capture, at least among the flute examples where there was a clear correlation between the ratings and objective features for phrasing as well as between the ratings and objective features for tone production. These correlations were not evident for the viol examples, either because the objective features are not able to capture the expressive mechanisms specific to violists or because there was less clear consensus among the listener ratings.

8.2 Further work

In order to support the conclusion that room acoustics affect the ability to perform in a historical baroque playing style, and that the one example of a baroque-era hall benefited such efforts while the modern example did not, a broader study could be undertaken which expands the number and variety of halls used. Evidence indicates that the Salon des Nobles was a fairly typical baroque hall for soloists and smaller ensembles (see section 3.3.3). However, a more comprehensive cataloguing of baroque-era performance spaces could assist in future efforts. An acoustic measurement campaign of smaller baroque-era halls may elucidate the acoustic commonalities between such halls, and also what sets them apart, acoustically, from other halls. Additionally, extending this research to ensembles would be useful to better separate the

acoustic needs of baroque soloists from the acoustic needs of baroque musicians in general.

Some improvements could be made in future similar studies. For example, improving the direct-to-reverberant ratio of the recorded signal could facilitate performance analysis. Increasing the variety of acoustics studied would also be beneficial. Even if the acoustics are not commonly encountered in real performance situations (such as an anechoic condition), they could provide useful benchmarks of the extent to which acoustics can affect performance practice. This could be useful for contextualizing the results of other similar studies. As it has been shown that the instrument influences how musicians adapt to acoustics, increasing the variety of instruments studied is important to better understand the significance of the role of the instrument in such studies.

Because of the previously mentioned issues with the VAE, the EVAA system was not able to be fully evaluated and there are still some novel components which show promise. A series of simplified studies intended to evaluate each component individually would be beneficial before combining all of these novel components for another complex study.

One of the abilities of a well-calibrated EVAA system would be the capacity to change different perceptual modalities individually. For instance, being able to use the acoustics of one room with the visual appearance of another would be useful to study the interaction of different modalities. Further improvements to the EVAA system could make such studies possible which are not possible in the real world.

Lastly, the efforts towards analyzing music performance, especially the baroque analysis framework, are promising steps towards more semantically meaningful but still somewhat automated evaluation. Further work could improve this analysis framework, such as incorporating more advanced multilevel models that take into consideration a multitude of features at once. Additionally, more and better-annotated data would be helpful in refining these analysis tools. For example, a set of recordings of musicians performing the same compositions multiple times but deviating intentionally among a set of specified expressive dimensions, as done in Fabian, Schubert, and Pulley (2010), would be essential in developing a robust music performance analysis framework.

Bibliography

Allen, J. B. and Berkley, D. A. (Apr. 1979). "Image method for efficiently simulating small-room acoustics." In: *Journal of the Acoustical Society of America* 65.4, pp. 943–950.

Amengual Gari, S. (Sept. 2017). "Investigations on the Influence of Acoustics on Live Music Performance using Virtual Acoustic Methods." PhD thesis. Detmold University of Music.

Amengual Gari, S., Kob, M., and Lokki, T. (Sept. 2019). "Analysis of trumpet performance adjustments due to room acoustics." In: *Proceedings of the International Symposium on Room Acoustics (ISRA 2019)*. Amsterdam, pp. 65–73.

Arend, J., Lübeck, T., and Pörschmann, C. (Mar. 2019). "A Reactive Virtual Acoustic Environment for Interactive Immersive Audio." In: *Proceedings of the AES International Conference on Immersive and Interactive Audio*. York.

Aucouturier, J.-J. and Bigand, E. (Oct. 2012). "Mel Cepstrum & Ann Ova: The difficult dialog between MIR and music cognition." In: *Proceedings of the 13th International Conference on Music Information Retrieval (ISMIR)*. Porto, pp. 397–402.

Barfield, W., Zeltzer, D., Sheridan, T., and Slater, M. (June 1995). "Presence and performance within virtual environments." In: *Virtual environments and advanced interface design*. Oxford University Press, pp. 473–513.

Baumont, O. (2007). *La musique à Versailles*. Actes sud/Château de Versailles/Centre de musique baroque de Versailles.

Békésy, G. von (1968). "Feedback phenomena between the stringed instrument and the musician." In: *Rockefeller University Review* 6, pp. 1–9.

Ben Hagai, I., Pollow, M., Vorländer, M., and Rafaely, B. (Oct. 2011). "Acoustic centering of sources measured by surrounding spherical microphone arrays." In: *Journal of the Acoustical Society of America* 130.4, pp. 2003–2015.

Beranek, L. (1962). *Music, Acoustics & Architecture*. New York London: John Wiley & Sons, Inc.

Beranek, L. (July 1992). "Concert hall acoustics—1992." In: *Journal of the Acoustical Society of America* 92.1, pp. 1–39.

Bittner, R. M., Salamon, J., Bosch, J. J., and Bello, J. P. (June 2017). "Pitch Contours as a Mid-Level Representation for Music Informatics." In: *Proceedings of the AES International Conference on Semantic Audio*. Erlangen.

Blauert, J. (1997). *Spatial Heading: The Psychophysics of Human Sound Localization*. Cambridge: MIT Press.

Bolzinger, S. and Risset, J. (1992). "A Preliminary Study on the Influence of Room Acoustics on Piano Performance." In: *Journal de Physique IV* 02, pp. C1–93 – C1–96.

Bolzinger, S., Warusfel, O., and Kahle, E. (1994). "Study of the influence of room acoustics on piano performance." In: *Journal de Physique Colloque C5*. Vol. 4, pp. C5–617 – C5–620.

Boren, B. (2019). "Computational acoustic musicology." In: *Digital Scholarship in the Humanities* 34.4, pp. 707–715.

Boren, B., Abraham, D., Naressi, R., Grzyb, E., Lane, B., and Merceruio, D. (Sept. 2019). "Acoustic simulation of Bach's performing forces in the Thomaskirche." In: *EAA Spatial Audio Signal Processing Symposium*. Paris, pp. 143–148.

Brereton, J. S. (Aug. 2014). "Singing in Space(s): Singing performance in real and virtual acoustic environments—Singers' evaluation, performance analysis and listeners' perception." PhD Dissertation. University of York.

Brown, J. C. and Vaughn, K. V. (Sept. 1996). "Pitch center of stringed instrument vibrato tones." In: *Journal of the Acoustical Society of America* 100.3, pp. 1728–1735.

Burred, J. J. and Lerch, A. (Sept. 2003). "A Hierarchical Approach To Automatic Musical Genre Classification." In: *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*. London.

Butt, J. (2002). *Playing With History: The Historical Approach to Musical Performance*. Cambridge University Press.

Cambouropoulos, E., Dixon, S., Goebl, W., and Widmer, G. (Aug. 2001). "Human Preferences for Tempo Smoothness." In: *Proceedings of the VII International Symposium on Systematic and Comparative Musicology*, pp. 18–26.

Canfield-Dafilou, E. K., Callery, E. F., Abel, J. S., and Berger, J. J. (Mar. 2019). "A Method for Studying Interactions between Music Performance and Rooms with Real-Time Virtual Acoustics." In: *Proceedings of the 146th AES Convention*. Dublin.

Chiang, W., Chen, S.-t., and Huang, C.-t. (2003). "Subjective Assessment of Stage Acoustics for Solo and Chamber Music Performances." In: *Acta Acustica United with Acustica* 89.5, pp. 848–856.

Chuan, C.-H. and Chew, E. (Sept. 2007). "A Dynamic Programming Approach To The Extraction Of Phrase Boundaries From Tempo Variations In Expressive Performances." In: *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*. Vienna, pp. 305–308.

Ciaramitaro, V. M., Chow, H. M., and Eglington, L. G. (Mar. 2017). "Cross-modal attention influences auditory contrast sensitivity: Decreasing visual load improves auditory thresholds for amplitude- and frequency-modulated sounds." In: *Journal of Vision* 17.3, p. 20.

Dalenbäck, B.-I. (May 2010). *Engineering principles and techniques in room acoustics prediction*. URL: https://www.catt.se/BNAM-Bergen-2010-CATT.pdf.

Dalenbäck, B.-I. (Feb. 2016). *CATT-Acoustic v9.1 with TUCT v2.0*.

Dammerud, J. J. (Sept. 2009). "Stage Acoustics for Symphony Orchestras in Concert Halls." PhD Dissertation. University of Bath.

Dart, T. (1954). *The Interpretation of Music*. London: Hutchinson's University Library.

Dixon, S. and Widmer, G. (Sept. 2005). "Match: A Music Alignment Tool Chest." In: *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*. London, pp. 492–497.

Domínguez, D. (2008). "A study of musicians' room acoustics conditions." MA thesis. Universidad Técnica de Dinamarca.

Donington, R. (1963). *The Interpretation of Early Music*. London: Faber and Faber.

Donington, R. (1973). *A Performer's Guide to Baroque Music*. New York: Charles Scribner's Sons.

Donington, R. (1982). *Baroque Music Style and Performance: A Handbook*. W.W. Norton & Company.

Eerola, T. and Toiviainen, P. (2004a). *MIDI Toolbox: MATLAB Tools for Music Research*. Jyväskylä: University of Jyväskylä: Kopijyvä.

Eerola, T. and Toiviainen, P. (Oct. 2004b). "MIR In Matlab: The MIDI Toolbox." In: *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*. Barcelona.

Everest, F. A. and Pohlmann, K. C. (2009). *Master handbook of acoustics*. 5th ed. OCLC: 434959408. New York: McGraw-Hill. ISBN: 978-0-07-160333-1.

Fabian, D. (2003). *Bach Performance Practice, 1945–1975: A Comprehensive Review of Sound Recordings and Literature*. 1st ed. Routledge.

Fabian, D. and Schubert, E. (Jan. 2009). "Baroque expressiveness and stylishness in three recordings of the D minor Sarabanda for solo violin by J.S. Bach." In: *Music Performance Research* 3, pp. 36–56.

Fabian, D., Schubert, E., and Pulley, R. (July 2010). "A Baroque Träumerei: The Performance and Perception of two Violin Renditions." In: *Musicology Australia* 32.1, pp. 27–44.

Farina, A. (May 2007). "Advancements in impulse response measurements by sine sweeps." In: *Proceedings of the 122nd AES Convention*. Vienna.

Fischinger, T., Frieler, K., and Louhivuori, J. (2015). "Influence of virtual room acoustics on choir singing." In: *Psychomusicology: Music, Mind, and Brain* 25.3, pp. 208–218.

Fitzgerald, D. (Sept. 2010). "Harmonic/Percussive Separation using Median Filtering." In: *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*. Graz.

Foteinou, A., Murphy, D. T., and Masinton, A. (May 2010). "Verification of Geometric Acoustics Based Auralization Using Room Acoustics Measurement Techniques." In: *Proceedings of the 128th AES Convention*. London.

Fragoulis, D., Papaodysseus, C., Exarhos, M., Roussopoulos, G., Panagopoulos, T., and Kamarotos, D. (May 2006). "Automated classification of piano-guitar notes." In: *IEEE Transactions on Audio, Speech and Language Processing* 14.3, pp. 1040–1050.

Frank, M. and Brandner, M. (Nov. 2019). "Perceptual Evaluation of Spatial Resolution in Directivity Patterns 2: coincident source/listener positions." In: *Fortschritte der Akustik: Tagungsband der 45 (DAGA 2019)*. Publisher: TU Ilmenau. Rostock, pp. 74–77.

Gade, A. C. (1989a). "Investigations of musician's room acoustics conditions in concert halls, Part II: Field experiments and synthesis of results." In: *Acta Acustica United with Acustica* 69.4, pp. 249–262.

Gade, A. C. (1989b). "Investigations of musicians' room acoustic conditions in concert halls, Part I: Methods and laboratory experiments." In: *Acta Acustica united with Acustica* 69.3, pp. 193–203.

Gade, A. C. (2007). "Acoustics in Halls for Speech and Music." In: *Springer Handbook of Acoustics*. Ed. by T. Rossing. New York: Springer Science+Business Media, pp. 301–350.

Gade, A. C. (Aug. 2010). "Acoustics for Symphony Orchestras; status after three decades of experimental research." In: *Proceedings of the International Symposium on Room Acoustics (ISRA 2010)*. Melbourne.

Galamian, I. (1962). *Principles of violin playing & teaching*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Gerzon, M. A. (Feb. 1973). "Periphony: With-Height Sound Reproduction." In: *Journal of the Audio Engineering Society* 21.1, pp. 2–10.

Guthrie, A. (2014). "Stage acoustics for musicians: A multidimensional approach using 3D ambisonic technology." PhD Dissertation. Rensselaer Polytechnic Institute.

Hamilton, B. and Bilbao, S. (Apr. 2018). "Wave-Based Room Acoustics Modelling: Recent Progress and Future Outlooks." In: *Proceedings of the Institute of Acoustics*. Vol. 40. Hamburg, pp. 160–161.

Haynes, B. (2007). *The End of Early Music: A Period Performer's History of Music*. Oxford University Press.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (Mar. 2005). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal inten-

sity, and fundamental frequency." In: *Journal of the Acoustical Society of America* 117.3, pp. 1417–1430.

Herrera, P. and Bonada, J. (1998). "Vibrato Extraction and Parameterization in the Spectral Modeling Synthesis Framework." In: *Proceedings of the 1st International Conference on Digital Audio Effects (DAFx-98)*. Barcelona.

Hindemith, P. (1952). *Johann Sebastian Bach: Heritage and Obligation*. New Haven: Yale University Press.

Hinton, G. E. and Roweis, S. (2002). "Stochastic Neighbor Embedding." In: *Advances in Neural Information Processing Systems*. Ed. by S. Becker, S. Thrun, and K. Obermayer. Vol. 15. MIT Press.

Houle, G. (1987). *Meter in music, 1600 - 1800*. Bloomington: Indiana University Press.

ISO (2009). *ISO 3382-1(E) Acoustics — Measurement of room acoustic parameters — Part 1: Performance spaces*.

ISO (June 2017). *ISO 532-1: Acoustics — Methods for calculating loudness — Part 1: Zwicker method*.

ITU-R (1998). *Recommendation ITU-R BS.1387-1, Method for objective measurements of perceived audio quality*.

Kato, K., Ueno, K., and Kawai, K. (May 2008). "Musicians' Adjustment of Performance to Room Acoustics, Part III: Understanding the Variations in Musical Expressions." In: *Journal of the Acoustical Society of America* 123.5, pp. 3610–3610.

Kato, K., Ueno, K., and Kawai, K. (July 2015). "Effect of Room Acoustics on Musicians' Performance. Part II: Audio Analysis of the Variations in Performed Sound Signals." In: *Acta Acustica united with Acustica* 101.4, pp. 743–759.

Katz, B. F. G., Leconte, S., and Stitt, P. (Sept. 2019). "EVAA: A platform for experimental virtual archeological-acoustics to study the influence of performance space." In: *Proceedings of the International Symposium on Room Acoustics (ISRA 2019)*. Amsterdam.

Katz, B. F. G., Poirier-Quinot, D., and Postma, B. N. J. (Sept. 2019). "Virtual reconstructions of the Théâtre de l'Athénée for archeoacoustic study." In: *Proceedings of the 23rd International Congress on Acoustics*. Aachen, Germany, pp. 303–310.

Kawai, K., Kato, K., Ueno, K., and Sakuma, T. (June 2013). "Experiment on adjustment of piano performance to room acoustics: Analysis of performance coded into MIDI data." In: *Proceedings of the International Symposium on Room Acoustics (ISRA 2013)*. Toronto.

Kearney, G. (Mar. 2010). "Auditory Scene Synthesis using Virtual Acoustic Recording and Reproduction." PhD Dissertation. Trinity College Dublin.

Kendall, R. and Carterette, E. (1990). "The Communication of Musical Expression." In: *Music Perception* 8.2, pp. 129–164.

Kim, K., Rosenthal, Z., Zielinski, D., and Brady, R. (Mar. 2012). "Comparison of Desktop, Head Mounted Display, and Six Wall Fully Immersive Systems using a Stressful Task." In: *Proceedings of the 2012 IEEE Virtual Reality Conference*. Costa Mesa, pp. 143–144.

Kirchhoff, H. and Lerch, A. (Mar. 2011). "Evaluation of Features for Audio-to-Audio Alignment." In: *Journal of New Music Research* 40.1, pp. 27–41.

Kleiner, M., Dalenbäck, B.-I., and Svensson, P. (1993). "Auralization-An Overview." In: *Journal of the Audio Engineering Society* 41.11, pp. 861–875.

Knight, T., Upham, F., and Fujinaga, I. (2011). "The potential for automatic assessment of trumpet tone quality." In: *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*. Miami, pp. 573–578.

Kob, M., Amengual Gari, S., and Schärer Kalkandjiev, Z. (2020). "Room Effect on Musicians' Performance." In: *The Technology of Binaural Understanding*. Ed. by J. Blauert and J. Braasch. Springer International Publishing, pp. 223–249.

Krokstad, A., Strom, S., and Sørsdal, S. (July 1968). "Calculating the acoustical room response by the use of a ray tracing technique." In: *Journal of Sound and Vibration* 8.1, pp. 118–125.

Kuhl, W. (Jan. 1954). "Über Versuche zur Ermittlung der günstigsten Nachhallzeit grosser Musikstudios." In: *Acta Acustica united with Acustica* 4.5, pp. 618–634.

Laird, I., Chapman, P., and Murphy, D. (2012). "Energy-based calibration of Virtual Performance Systems." In: *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*. York.

Laird, I., Murphy, D., Chapman, P., and Jouan, S. (May 2011). "Development of a Virtual Performance Studio with Application of Virtual Acoustic Recording Methods." In: *Proceedings of the 130th AES Convention*. London.

Lashley, K. (1951). "The problem of serial order in behavior." In: *Cerebral Mechanisms in Behavior: The Hixon Symposium*. Ed. by L. Jeffress. New York: Wiley, pp. 112–136.

Lawson, C. (2002). "Performing through history." In: *Musical Performance, A Guide to Understanding*. Ed. by J. Rink. New York: Cambridge University Press, pp. 3–16.

Lawson, C. and Stowell, R. (2003). *The historical performance of music: an introduction*. Cambridge handbooks to the historical performance of music. Cambridge University Press.

Lawson, C. and Stowell, R., eds. (Feb. 2012). *The Cambridge History of Musical Performance*. 1st ed. Cambridge University Press.

Lerch, A. (Aug. 2008). "Software-Based Extraction of Objective Parameters from Music Performances." PhD Dissertation. Technischen Universität Berlin.

Lerch, A. (2012). *Audio content analysis: an introduction*. Hoboken, N.J: Wiley.

Lerch, A., Arthur, C., Pati, A., and Gururani, S. (Nov. 2020). "An Interdisciplinary Review of Music Performance Analysis." In: *Transactions of the International Society for Music Information Retrieval* 3.1, pp. 221–245. ISSN: 2514-3298.

Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (June 2001). "A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory." In: *Presence* 10, pp. 282–297.

Li, S., Dixon, S., and Plumbley, M. D. (Oct. 2017). "Clustering Expressive Timing With Regressed Polynomial Coefficients Demonstrated By A Model Selection Test." In: *Proceedings of the 18th International Conference on Music Information Retrieval (ISMIR)*. Suzhou, pp. 457–463.

Lobdell, B. E. and Allen, J. B. (Jan. 2007). "A model of the VU (volume-unit) meter, with speech applications." In: *Journal of the Acoustical Society of America* 121.1, pp. 279–285.

Logan, B. (Oct. 2000). "Mel Frequency Cepstral Coefficients For Music Modeling." In: *Proceedings of the 1st International Conference on Music Information Retrieval (ISMIR)*. Plymouth.

Lokki, T. and Pätynen, J. (Dec. 2020). "Absorption of gilding in concert halls." In: *Proceedings of the Forum Acusticum (FA2020)*. Lyon, pp. 1727–1730.

Luizard, P., Brauer, E., and Weinzierl, S. (Mar. 2019). "Singing in physical and virtual environments: how performers adapt to room acoustical conditions." In: *Proceedings of the AES International Conference on Immersive and Interactive Audio*. York.

Luizard, P. and Henrich Bernardoni, N. (July 2020). "Changes in the voice production of solo singers across concert halls." In: *Journal of the Acoustical Society of America* 148.1, EL33–EL39.

Luizard, P., Steffens, J., and Weinzierl, S. (Feb. 2020). "Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions?" In: *Journal of the Acoustical Society of America* 147.2, EL132–EL137.

Maaten, L. v. d. and Hinton, G. (2008). "Visualizing Data using t-SNE." In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605.

Mahalanobis, P. C. (Apr. 1936). "On the generalized distance in statistics." In: *Proceedings of the National Institute of Sciences of India* 2.1, pp. 49–55.

Marshall, A. H., Gottlob, D., and Alrutz, H. (Nov. 1978). "Acoustical conditions preferred for ensemble." In: *Journal of the Acoustical Society of America* 64.5, pp. 1437–1442.

Mauch, M. and Dixon, S. (May 2014). "PYIN: A fundamental frequency estimator using probabilistic threshold distributions." In: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence, Italy: IEEE, pp. 659–663.

McAdams, S. (2013). "Musical Timbre Perception." In: *The Psychology of Music*. Elsevier, pp. 35–67.

McAdams, S., Cunible, J.-c., Carlyon, R. P., Darwin, C. J., and Russell, I. J. (June 1992). "Perception of timbral analogies." In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 336.1278, pp. 383–389.

Meyer, J. (2009). *Acoustics and the Performance of Music, Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*. Trans. by U. Hansen. Fifth Edition. Springer Science+Business Media.

O'Shaughnessy, D. (1987). *Speech Communication: Human and Machine*. Addison-Wesley series in electrical engineering. Addison-Wesley Publishing Company.

Otondo, F. and Rindel, J. H. (2004). "The Influence of the Directivity of Musical Instruments in a Room." In: *Acta Acustica united with Acustica* 90.6, pp. 1178–1184.

Palmer, C. (1989). "Mapping Musical Thought to Musical Performance." In: *Journal of Experimental Psychology: Human Perception and Performance* 15.12, pp. 331–346.

Palmer, C. (1997). "Music Performance." In: *Annual Review of Psychology* 48, pp. 115–138.

Pang, H.-S. and Yoon, D.-H. (July 2005). "Automatic detection of vibrato in monophonic music." In: *Pattern Recognition* 38.7, pp. 1135–1138.

Panton, L., Yadav, M., Cabrera, D., and Holloway, D. (June 2019). "Chamber musicians' acoustic impressions of auditorium stages: Relation to spatial distribution of early reflections and other parameters." In: *Journal of the Acoustical Society of America* 145.6, pp. 3715–3726.

Pietrzyk, A. and Kleiner, M. (Mar. 1997). "The Application of the Finite-Element Method to the Prediction of Soundfields of Small Rooms at Low Frequencies." In: *Proceedings of the 102nd AES Convention*. Munich.

Ponsford, D. (2012). "Instrumental performance in the seventeenth century." In: *The Cambridge History of Musical Performance*. Ed. by C. Lawson and R. Stowell. Cambridge University Press, pp. 421–447.

Postma, B. N. J., Demontis, H., and Katz, B. F. G. (Jan. 2017). "Subjective Evaluation of Dynamic Voice Directivity for Auralizations." In: *Acta Acustica United with Acustica* 103.2, pp. 181–184.

Postma, B. N. J., Dubouilh, S., and Katz, B. F. G. (Apr. 2019). "An archeoacoustic study of the history of the Palais du Trocadero (1878–1937)." In: *Journal of the Acoustical Society of America* 145.4, pp. 2810–2821.

Postma, B. N. J. and Katz, B. F. G. (Sept. 2015). "Creation and calibration method of acoustical models for historic virtual reality auralizations." In: *Virtual Reality* 19, pp. 161–180.

Postma, B. N. J. and Katz, B. F. G. (Mar. 2016). "Dynamic voice directivity in room acoustic auralizations." In: *Fortschritte der Akustik: Tagungsband der 42 (DAGA 2016)*. Aachen, pp. 352–355.

Postma, B. N. J. and Katz, B. F. G. (Nov. 2017). "The influence of visual distance on the room-acoustic experience of auralizations." In: *Journal of the Acoustical Society of America* 142.5, pp. 3035–3046.

Postma, B. N. J., Poirier-Quinot, D., Meyer, J., and Katz, B. F. G. (June 2016). "Virtual Reality Performance Auralization in a Calibrated Model of Notre-Dame Cathedral." In: *EAA EuroRegio2016*. Porto, Portugal.

Postma, B. N. J., Tallon, A., and Katz, B. F. G. (2015). "Calibrated Auralization Simulation of the Abbey of Saint-Germain-Des-Prés for Historical Study." In: *Proceedings of the Institute of Acoustics*. Vol. 37 Pt. 3, pp. 190–197.

Quantz, J. J. (1752). *Essai d'une Methode pour Apprendre à Jouer de la Flute Traversiere (On Playing the Flute)*. Trans. by E. R. Reilly. Faber & Faber 1966. Voss.

Regnier, L. and Peeters, G. (Apr. 2009). "Singing Voice Detection in Music Tracks using Direct Voice Vibrato Detection." In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Taipei, pp. 1685–1688.

Repp, B. H. (Aug. 1998). "A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major." In: *The Journal of the Acoustical Society of America* 104.2, pp. 1085–1100.

Repp, B. H. (July 1999a). "A microcosm of musical expression. III. Contributions of timing and dynamics to the aesthetic impression of pianists' performances of the initial measures of Chopin's Etude in E Major." In: *Journal of the Acoustical Society of America* 106.1, pp. 469–478.

Repp, B. H. (July 1999b). "Effects of Auditory Feedback Deprivation on Expressive Piano Performance." In: *Music Perception* 16.4, pp. 409–438. ISSN: 0730-7829.

Rink, J. (2002). *Musical performance: a guide to understanding*. Cambridge; New York: Cambridge University Press.

Rothstein, J. (1995). *MIDI: A comprehensive introduction*. 2nd. Madison, Wisconsin: A-R Editions Inc.

Sabine, W. C. (1964). *Collected papers on acoustics*. New York: Dover Publications.

Salamon, J. and Gomez, E. (2012). "Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics." In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.6, pp. 1759–1770.

Salmon, F., Hendrickx, É., Épain, N., and Paquier, M. (Sept. 2020). "The Influence of Vision on the Perceived Differences Between Sound Spaces." In: *Journal of the Audio Engineering Society* 68.7/8, pp. 522–531.

Savioja, L. and Svensson, U. P. (Aug. 2015). "Overview of geometrical room acoustic modeling techniques." In: *Journal of the Acoustical Society of America* 138.2, pp. 708–730.

Scaringella, N. and Zoia, G. (Sept. 2005). "On the Modeling of Time Information for Automatic Genre Recognition Systems in Audio Signals." In: *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*. London, pp. 666–671.

Schärer Kalkandjiev, Z. (2015). "The Influence of Room Acoustics on Solo Music Performances. An Empirical Investigation." PhD Dissertation. Technischen Universität Berlin.

Schärer Kalkandjiev, Z. and Weinzierl, S. (Sept. 2015). "Playing slow in reverberant rooms — examination of a common concept based on empirical data." In: *Proceedings of the 3rd Vienna Talk on Music Acoustics*. Vienna, pp. 215–219.

Schärer Kalkandjiev, Z. and Weinzierl, S. (2018). "An Instrument for Measuring the Perception of Room Acoustics from the Perspective of Musicians: The Stage Acoustic Quality Inventory (STAQI)." In: *Fortschritte der Akustik: Tagungsband der 44 (DAGA 2018)*, pp. 1747–1750.

Schiltz, K. (Feb. 2003). "Church and chamber: the influence of acoustics on musical composition and performance." In: *Early Music*, pp. 64–80.

Schoonderwaldt, E., Guettler, K., and Askenfelt, A. (Aug. 2003). "Effect of the Width of the Bow Hair on the Violin String Spectrum." In: *Proceedings of the Stockholm Music Acoustics Conference (SMAC)*. Stockholm, pp. 91–94.

Schreiber, H., Zalkow, F., and Müller, M. (Oct. 2020). "Modeling and estimating local tempo: A case study on Chopin's Mazurkas." In: *Proceedings of the 21st International Conference on Music Information Retrieval (ISMIR)*. Montréal, pp. 773–779.

Schroeder, M. R. (Mar. 1965). "New Method of Measuring Reverberation Time." In: *Journal of the Acoustical Society of America* 37.3, pp. 409–412.

Schubert, E. and Fabian, D. (Oct. 2006). "The dimensions of baroque music performance: a semantic differential study." In: *Psychology of Music* 34.4, pp. 573–587.

Schubert, E. and Fabian, D. (July 2014). "A taxonomy of listeners' judgments of expressiveness in music performance." In: *Expressiveness in music performance: Empirical approaches*

across styles and cultures. Ed. by D. Fabian, R. Timmers, and E. Schubert. Oxford University Press, pp. 283–303.

Schutte, M., Ewert, S. D., and Wiegrebe, L. (Mar. 2019). "The percept of reverberation is not affected by visual room impression in virtual environments." In: *Journal of the Acoustical Society of America* 145.3, EL229–EL235.

Seashore, C. E. (1938). *Psychology of Music*. New York and London: McGraw-Hill Book Company.

Shabtai, N., Behler, G., Vorländer, M., and Weinzierl, S. (Feb. 2017). "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments." In: *Journal of the Acoustical Society of America* 141.2, pp. 1246–1256.

Shabtai, N. R. and Vorländer, M. (Apr. 2015). "Acoustic centering of sources with high-order radiation patterns." In: *Journal of the Acoustical Society of America* 137.4, pp. 1947–1961.

Shiratuddin, M. F. and Sulbaran, T. (July 2006). "A comparison of virtual reality displays-suitability, details, dimensions and space." In: *Proceedings of the 9th International Conference on Engineering Education*. San Juan.

Siedenburg, K., Fujinaga, I., and McAdams, S. (Jan. 2016). "A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology." In: *Journal of New Music Research* 45.1, pp. 27–41.

Sloboda, J. A. (1982). "Music Performance." In: *Psychology of Music*. Ed. by D. Deutsch. New York: Academic Press, pp. 479–496.

Sloboda, J. A. (Oct. 2000). "Individual differences in music performance." In: *Trends in Cognitive Sciences* 4.10, pp. 397–403.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press.

Ternström, S. (1989). "Long-time average spectrum characteristics of different choirs in different rooms." In: *Quarterly Progress and Status Report* 30.3, pp. 15–31.

Thery, D., Poirier-Quinot, D., Postma, B. N. J., and Katz, B. F. G. (2017). "Impact of the Visual Rendering System on Subjective Auralization Assessment in VR." In: *Virtual Reality and Augmented Reality*. Ed. by J. Barbic, M. D'Cruz, M. E. Latoschik, M. Slater, and P. Bourdot. Vol. 10700. Cham: Springer International Publishing, pp. 105–118.

Timmers, R. (Jan. 2005). "Predicting the similarity between expressive performances of music from measurements of tempo and dynamics." In: *Journal of the Acoustical Society of America* 117.1, pp. 391–399.

Todd, N. P. M. (June 1992). "The dynamics of dynamics: A model of musical expression." In: *Journal of the Acoustical Society of America* 91.6, pp. 3540–3550.

Todd, N. P. M. (Apr. 1993). "Vestibular Feedback in Musical Performance: Response to "Somatosensory Feedback in Musical Performance" (Edited by Sundberg and Verrillo)." In: *Music Perception* 10.3, pp. 379–382.

Tredinnick, R., Boettcher, B., Smith, S., Solovy, S., and Ponto, K. (Mar. 2017). "Uni-CAVE: A Unity3D Plugin for Non-head Mounted VR Display Systems." In: *Proceedings of the 2017 IEEE Virtual Reality Conference*. Los Angeles.

Tzanetakis, G. and Cook, P. (July 2002). "Musical genre classification of audio signals." In: *IEEE Transactions on Speech and Audio Processing* 10.5, pp. 293–302.

Ueno, K., Kato, K., and Kawai, K. (May 2010). "Effect of Room Acoustics on Musicians' Performance. Part I: Experimental Investigation with a Conceptual Model." In: *Acta Acustica united with Acustica* 96.3, pp. 505–515.

Ueno, K., Yasuda, K., Tachibana, H., and Ono, T. (2001). "Sound field simulation for stage acoustics using 6-channel system." In: *Acoustical Science and Technology* 22.4, pp. 307–309.

Vassilantonopoulos, S. and Mourjopoulos, J. (Sept. 2001). "Virtual Acoustic Reconstruction of Ritual and Public Spaces of Ancient Greece." In: *Acta Acustica united with Acustica* 87.5, pp. 604–609.

Vigeant, M. C., Wang, L. M., and Rindel, J. H. (Sept. 2007). "Investigations of Multi-Channel Auralization Technique for Solo Instruments and Orchestra." In: *Proceedings of the 19th International Congress on Acoustics*. Madrid.

Walter, B. (1961). *Of Music and Music-Making*. Trans. by P. Hamburger. Faber & Faber.

Wang, L. M. and Vigeant, M. C. (2004). "Objective and subjective evaluation of the use of directional sound sources in auralizations." In: *Proceedings of the 18th International Congress on Acoustics*. Kyoto, pp. 2711–2714.

Wang, S., Ewert, S., and Dixon, S. (Nov. 2016). "Robust and Efficient Joint Alignment of Multiple Musical Performances." In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.11, pp. 2132–2145.

Weninger, F., Amir, N., Amir, O., Ronen, I., Eyben, F., and Schuller, B. (Mar. 2012). "Robust feature extraction for automatic recognition of vibrato singing in recorded polyphonic music." In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Kyoto, pp. 85–88.

Wilson, G. D. and Roland, D. (2002). "Performance Anxiety." In: *The Science & Psychology of Music Performance: Creative Strategies for Teaching and Learning*. Ed. by R. Parncutt and G. E. McPherson. New York: Oxford University Press, pp. 47–61.

Zotter, F. and Frank, M. (2012). "All-round ambisonic panning and decoding." In: *Journal of the Audio Engineering Society* 60.10, pp. 807–820.

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics facts and models*. Ed. by M. Schroeder. Berlin Heidelberg New York: Springer-Verlag.

Appendix A

Questionnaires

This appendix contains the full text of all questionnaires discussed in section 4.2 in english and in french.

Table A.1 Open-ended questions at end of acoustic rating questionnaire.

How does this performance space make you feel?	Pouvez-vous décrire comment cet espace a impacté votre jeu? Votre ressenti?
Was your performance influenced by the way you felt? If yes, in what way?	En quoi votre état émotionnel aujourd'hui a-t-il affecté la façon dont vous avez joué?
Were you conscious of any adjustments to your performance due to the space? And if yes, to what would you attribute this?	Avez-vous été conscient des ajustements de votre jeu vis à vis de cet espace? Et si oui, à quoi êtes-vous ajusté?

Table A.2 First half of acoustic rating questionnaire.

Amount of reverberation	Force de la réverbération
Little — A lot	Faible — Forte
Timbre	Timbre
Displeasing — Pleasing	Rugueux — Fluide
Reverberance	Réverbérance
Dry — Reverberant	Sec — Réverbérant
Echoes/disturbing reflections	Echos/réflexions gênantes
Many disturbing reflections — None	Beaucoup — Aucun
Studio-like — Church-like	Sonne comme un studio — Sonne comme une église
Ease of maintaining tempo	Facilité à maintenir le tempo
Easy — Difficult	Facile — Difficile
Resonance	Résonance
Little — A lot	Pas résonant — Très résonant
Suitability	Adaptation de la salle au jeu
Not suitable — Suitable	Ne convient pas — Est adaptée
Ease of hearing self	Facilité à s'entendre
Difficult — Easy	Difficile — Facile

Table A.3 Second half of acoustic rating questionnaire.

Sense of envelopment		Enveloppement	
Very frontal	Very enveloping	Peu enveloppant	Très enveloppant
Projection		Projection	
Does not carry	Carries	Ne porte pas	Porte bien
Ease of hearing dynamics		Facilité à entendre la dynamique du jeu	
Easy	Difficult	Facile	Difficile
Duration of reverberation		Durée de la réverbération	
Short	Long	Courte	Longue
Comfort		Confort sonore	
Uncomfortable	Comfortable	Inconfortable	Confortable
Room response		Réponse de la salle	
Dead	Live	Sèche	Vivante
Transparency		Transparence/clarté de jeu	
Muddy	Clear	Brouillé	Clair
Enjoyment		Plaisir du jeu	
Not enjoyable	Enjoyable	Déplaisant	Plaisant
Quality		Qualité acoustique	
Bad acoustics	Good acoustics	Mauvaise	Bonne

Table A.4 Virtual questionnaire ratings questions.

What I saw in the virtual environment was realistic/natural Strongly disagree — Strongly agree	Ce que j'ai VU dans l'environnement virtuel était réaliste/naturel Pas du tout d'accord — Tout à fait d'accord
What I HEARD in the virtual environment was realistic/natural Strongly disagree — Strongly agree	Ce que j'ai ENTENDU dans l'environnement virtuel était réaliste/naturel Pas du tout d'accord — Tout à fait d'accord
I perceived a disconnect between my actions and the visual virtual environment Strongly disagree — Strongly agree	J'ai perçu une décalage entre mes actions et l'environnement virtuel visuelle Pas du tout d'accord — Tout à fait d'accord
I perceived a disconnect between my actions and the virtual sound environment Strongly disagree — Strongly agree	J'ai perçu une décalage entre mes actions et l'environnement virtuel sonore Pas du tout d'accord — Tout à fait d'accord
The balance of the sound of my instrument with the reverb of the room Weak — Strong	L'équilibre du son de mon instrument par rapport à la réverbération de la salle Faible — Forte
What I SAW and HEARD was consistent Strongly disagree — Strongly agree	Ce que j'ai VU et ENTENDU était cohérent Pas du tout d'accord — Tout à fait d'accord
The force of reverberation in relation to my instrument Weak — Strong	La force de la réverbération par rapport a mon instrument Faible — Forte
Compared to my movements, what I HEARD in the virtual environment was realistic/natural Strongly disagree — Strongly agree	Par rapport à mes mouvements, ce que j'ai ENTENDU dans l'environnement virtuel était réaliste/naturel Pas du tout d'accord — Tout à fait d'accord
Compared to my movements, what I saw in the virtual environment was realistic/natural Strongly disagree — Strongly agree	Par rapport à mes mouvements, ce que j'ai VU dans l'environnement virtuel était réaliste/naturel Pas du tout d'accord — Tout à fait d'accord
The coloring/timbre of the virtual acoustics was realistic/natural Strongly disagree — Strongly agree	La coloration/timbre de l'acoustique virtuelle était réaliste/naturelle Pas du tout d'accord — Tout à fait d'accord
The acoustics seem to me faithful to the experience of really playing in the room Strongly disagree — Strongly agree	CL'acoustique me semble fidèle à l'expérience de jeu réel dans la salle Pas du tout d'accord — Tout à fait d'accord

Table A.5 Open-ended questions at end of virtual rating questionnaire.

Has being in a virtual environment had an impact on your performance? If so, how?	Le fait d'être dans un environnement virtuel a-t-il eu un impact sur votre performance? Si oui, comment?
Do you think the virtual acoustic environment can be improved? If so, how?	Pensez-vous que l'environnement acoustique virtuel puisse être amélioré? Si oui, comment?

Table A.6 Comparison questionnaire.

How do you rate the difference between your playing experiences (musical performance) in the two real rooms (Versailles versus La Cité de la Musique)?	Comment évaluez-vous la différence entre vos expériences de jeu (performance musicale) dans les deux salles réelles (Versailles versus La Cité de la Musique) ?
Very different Similar	Très différentes Similaires
How do you assess the difference between your playing experiences (musical performance) in both contexts (Real rooms versus Virtual Clones)	Comment évaluez-vous la différence entre vos expériences de jeu (performance musicale) dans les deux contextes (Salles réelles versus Clones virtuels)
Very different Similar	Très différentes Similaires

Appendix B

Preliminary analysis of vocal ensemble performances in real-time historical auralizations of the Palais des Papes

The article below describes a study that was conducted as part of this thesis. Although it is not directly related to the main objectives of the thesis, it is included in the appendix for reference as it may still be relevant to some aspects of the subject matter.

PROCEEDINGS of the 2nd Symposium: The Acoustics of Ancient Theatres
6–8 July 2022 Verona, Italy

Preliminary analysis of vocal ensemble performances in real-time historical auralizations of the Palais des Papes

Julien De Muynke[1,2,*]; Nolan Eley[1,3,*]; Julien Ferrando[4,*]; Brian F.G. Katz[1,*]

[1] Institut Jean le Rond ∂'Alembert, Sorbonne Université/CNRS, France
[2] Eurecat, Centre Tecnològic de Catalunya, Multimedia Technologies Group, Spain
[3] ETIS Laboratory, CY Cergy Paris University, ENSEA, CNRS, France
[4] AMU-CNRS, PRISM, UMR 7061, Marseille, France
julien.de_muynke@sorbonne-universite.fr, nolan.eley@sorbonne-universite.fr, julien.ferrando@univ-amu.fr, brian.katz@sorbonne-universite.fr

ABSTRACT

In the middle of the 14th century, the recently constructed Great Clementine Chapel of the Palais des Papes had a flourishing reputation for the composition and interpretation of polyphonic singing in the emerging Ars Nova musical style. In modern times, the space is still employed for musical performances. However, the acoustic conditions between the two periods vary greatly, and as such, can be expected to have an impact on vocal performances. As part of the IMAPI and PHE projects, the impact of the acoustics of the Great Clementine Chapel on the performance of a conducted vocal ensemble specializing in medieval music was examined for these two periods. A numerical simulation of the medieval acoustics was developed, based on a calibrated geometrical acoustics model of the modern-day chapel which was then regressed in time to a historically informed medieval state. Experiments were carried out with singers performing repetitions of several pieces in a Virtual Acoustic Environment (VAE) using close-mics and headphone renderings. Recorded performances were analyzed using various metrics, with objective results paired with questionnaires acquired for each VAE condition. Preliminary analysis of these results is presented in this study.

1. INTRODUCTION

1.1 The Palais des Papes in the 14th century

In 1309, the Holy See relocated to Avignon, France, where Pope Clement V established his residence, remaining there until 1403 [1]. There, construction of the Palais des Papes, which is the largest Gothic edifice ever built, started in 1335 under the pontificate of Benoît XII, was continued by Pope Clement VI from 1342, and was completed in 1352. Masses accompanied by music performance, especially polyphonic singing, were usually performed in the Great Clementine Chapel (a.k.a. Great Chapel). The Great Chapel attracted music composers, cantors, and musicians, particularly those belonging to the movement known today as the Ars Nova style. Ars Nova is a polyphonic musical style that developed in France in the 14th century as the successor of the Ars Antiqua exemplified by the School of Notre-Dame. It allowed for a higher degree of musical expressiveness and for more elaborate rhythmic modes due to a new standardized system of musical notation, even though some studies have shown that interpreting Ars Antiqua and Ars Nova as two radically different styles is probably excessive [2].

1.2 The impact of room acoustics on musical performance

Practitioners of choral music have long been aware that room acoustics play a significant role in musical performance [3]. However, despite this awareness, there has been no unified approach or theory to guide performance practice in response to different acoustic environments. In fact, while empirical studies have shown measurable effects on musical performance as a result of changes in acoustics, these effects tend to be rather small and/or individual [4, 5]. In short, it is still not well known precisely how room acoustics affect musical performance, and the evidence within the context of historical music is even less sufficient.

This study aims at assessing the impact of room acoustics on the musical performance of the conducted vocal ensemble *Diabolus in Musica*, consisting of three male vocalists (one baritone and two tenors) trained in medieval performance methods and with familiarity singing in the modern-day Great Chapel of the Palais des Papes.

2. VIRTUAL ACOUSTICS OF PALAIS DES PAPES

The acoustics of the Great Chapel of the Palais des Papes in two historical states, namely medieval (ca. 1362) and modern (ca. 2020) states, were generated through geometrical acoustic (GA) models designed in CATT-Acoustic v9.1e.

For that, the geometry of the room was first designed with reference to a 3D laser scan point cloud. Then, the definition of the construction materials of the modern-day room was carried out through an acoustic calibration procedure [6], in which the acoustic properties of the materials were adjusted until various acoustic metrics fell within the range of ± 1 JND of the measured acoustic values, based on recorded room impulse responses.

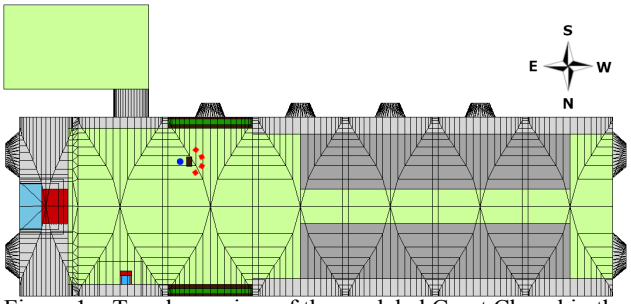


Figure 1 – Top-down view of the modeled Great Chapel in the medieval state. The various materials are represented by different colors. ■: singers; ●: conductor; ▮: lectern.

Next, a GA model of the medieval state of the Great Chapel was obtained by carrying out a time regression of the modern-day GA model with the help of historical records of the interior furnishing and decoration from that time era [7]. The presence of absorbing materials such as wall tapestries, floor rush matting, stalls, pews, the pope's throne, cloth on top of the altar and the stalls, and canopies above the altar and the pope's throne (represented by different colors in Fig. 1) make the medieval Great Chapel significantly less reverberant than the modern-day Great Chapel which is, essentially, a large empty shoe box made of dense limestone. A comparison of the reverberation time between the two historical states of the Great Chapel is shown in Fig. 2, including the presence of a simulated audience which tends to reduce the reverberation time.

3. SINGING EXPERIMENTS

The experiments were aimed at assessing the impact of the acoustics of the Great Chapel under different conditions on the musical performance of the vocal ensemble *Diabolus in Musica*. It was therefore important to provide the singers with performance conditions that were as close as possible to a real concert situation, while allowing for an audio recording quality sufficient to be used for further objective analysis (see Section 4). The experimental setup used in this study was guided by findings in [8].

3.1 Hardware setup

The singing experiments were carried out in a hemianechoic room at the PRISM laboratory in order to reduce interference of the recording room as much as possible. Singers were each equipped with a close microphone (headband cardioid microphone, DPA4088) in order to reduce the level of inter-singer crosstalk while having them distributed close together in a usual concert configuration. Virtual Acoustic Environments (VAEs) were reproduced for each individual singer over open-back headphones (Sennheiser HD650). Open-back headphones were chosen as they allow the direct sound from one's own voice and from other singers to pass through relatively unobstructed, while the reverberated voice sound is reproduced inside the headphones. The ratio between direct and reverberant sound levels was adjusted prior to the experiments through a calibration procedure to achieve realistic balance [9].

3.2 Auralization system

The VAEs were auralized via convolution with prerendered Binaural Room Impulse Responses (BRIRs) from the GA models with the direct sound removed. Each singer playing the role of a source and a receiver at the same time, the reproduction system included a total of $3 \times 3 = 9$ BRIRs for 3 singers, to which 3 extra BRIRs intended for the conductor were added, for a total of 12 BRIRs (i.e. 24 convolutions). The computational cost of such a reproduction system is of concern, especially considering that the BRIRs in the modern-day Great Chapel are 10s long in accordance with the reverberation time in the lowest octave band (see Fig. 2). The auralization architecture was created in MaxMSP to facilitate real-time processing. The convolution was done using the object `multiconvolve` from the HISS Impulse Response Toolbox[1] which employs a fixed partitioning scheme. The system was configured with an internal audio buffer size of 64 samples at 48 kHz, corresponding to an I/O delay of 1.3ms, without artifacts. This delay was compensated for by removing the 64 leading zeros in the BRIRs, providing correct time synchronization between the natural direct sound and the virtual reverberated sound.

3.3 The VAEs

The VAEs used in this study differ in their historical state, namely medieval (ca. 1362, when the Great Chapel was actively used for papal masses) and modern-day (the Great Chapel is still used as a performance space for concerts of vocal ensembles). The choir was positioned in the third bay starting from the east (the *parcus cantorum*, which included stalls in the medieval era), halfway between the chapel symmetry axis and the southern wall. The singers were distributed along an arc spanning 90° centered on the position of a virtual lectern and at a distance of 1.2 m.

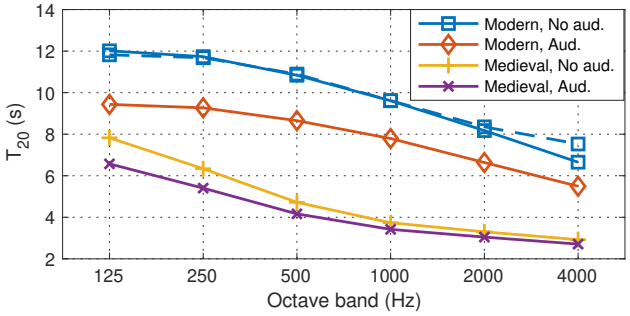


Figure 2 – Estimated T_{20} as a function of frequency band, averaged across sixteen positions evenly distributed along the Great Chapel. For the modern, unoccupied state, the measured (– –) and modeled (—) T_{20} s are shown.

[1] http://eprints.hud.ac.uk/id/eprint/14897/

Sources were simulated with the singing voice directivity pattern `singer.SD1` from CATT-Acoustic, pointing at the virtual lectern with the conductor facing them. The positions of the singers and conductor are shown in Fig. 1.

3.4 Experiment protocol

The repertoire comprised two pieces of polyphonic music from the Ars Nova style: "Petre Clemens" by Philippe de Vitry, and "Kyrie Rex Angelorum" (anonymous). The recordings were organized in separate sessions, each rendering either the medieval or modern-day room. In each session, the music pieces were interleaved and repeated 3 times. To compensate for a slightly different positioning of each individual singer's microphone, each microphone gain was adjusted at the beginning of each session to ensure a good consistency between individual singers' voice levels in the overall rendered audio scene. After each session, the participants answered a questionnaire on their subjective experience regarding the simulated acoustics and their performance in that particular space.

4. MUSIC PERFORMANCE ANALYSIS

Features were extracted from recordings of the performances which can be broken down into four musical categories: timing, dynamics, timbre, and pitch.

To represent timing, the note-level tempo was calculated by taking the inverse of the time interval between the onsets of adjacent notes weighted by the written note duration. The note onsets were obtained by manual annotation of one performance followed by audio-to-audio alignment using the `Match` plug-in[2] in Sonic Visualizer[3] followed by manual verification and adjustment.

A-weighted RMS was chosen to serve as a measure of musical dynamics or loudness. As a simplified measure of timbre, the spectral centroid was calculated. The spectral centroid represents the center of gravity of the spectrum and has been shown to be strongly correlated with the perception of a signal's "brightness" [10]. Both the spectral centroid and A-weighted RMS were extracted as time-series vectors with a window size of 2048 samples and a hop size of 10 ms. These vectors were later shortened to a grid of 8th note durations utilizing the note onset information necessary for calculating tempo. The 8th note segments corresponding to rests in the score were removed prior to analysis.

The fundamental frequency of each singer's performance, which was necessary to calculate higher level pitch-related features, was extracted using the `pYin` algorithm [11] in Sonic Visualizer. Because vibrato is a common expressive tool for singers, both the vibrato rate (mean pitch variation rate, in Hz) and vibrato extent (mean absolute distance to the note pitch center, in cents) were calculated on all notes with a duration $\geq$ a dotted-quarter.

Other researchers have used pitch drift, or the amount the pitch center changes throughout the course of the piece, as an indicator of overall ensemble intonation [5]. However, some amount of pitch drift is normal and may simply be the result of an unaccompanied ensemble singing in a non-equal temperament [12]. So, rather than using pitch drift as a measure of intonation, normalized pitch error was used. Normalized pitch error describes individual note intonation compared to its nominal pitch adjusted for overall pitch drift; a slight modification of the methodology outlined in [5].

5. RESULTS

As a preliminary analysis of the data, box plots were produced for each feature to examine whether or not there was a significant difference between the two acoustic settings. No significant differences were found with the exception of the loudness feature which indicated that each singer sang louder overall in the modern acoustics (see Fig. 3), however, the average difference was only $1.2\,\text{dB} \pm 0.2\,\text{dB}$ in "Kyrie Rex Angelorum" and $0.8\,\text{dB} \pm 0.5\,\text{dB}$ in "Petre Clemens". A Friedman test with singers as blocks and acoustics as group variable showed that these differences were statistically significant ($p < 0.001$ for both pieces). This greater vocal effort may be partly as compensation for the more reverberant nature of the modern acoustics, however, given that the difference is so small, too much emphasis should not be put on this finding at this time.

Rating questionnaires were given to the participants which asked about the following categories: reverberation, ease of ensemble singing, sound support, quality of the space, and size. No broad consensus was reached in any of these categories with the exception of reverberation, in which all the participants correctly ascertained that the modern state was more reverberant than the medieval state.

In addition to the questionnaire, participants were also encouraged to provide commentary freely which indicated

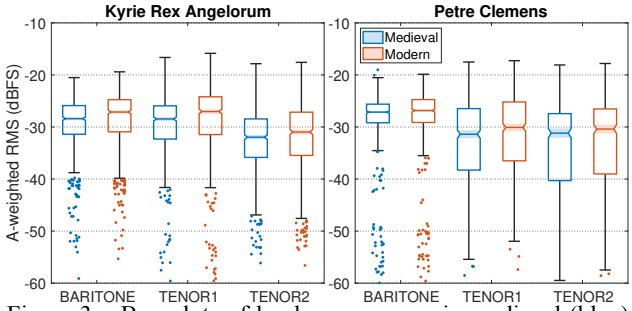


Figure 3 – Box plots of loudness measure in medieval (blue) and modern (red) acoustics by voice, for "Kyrie Rex Angelorum" (left) and "Petre Clemens" (right). Whiskers extend from upper & lower quartiles to non-outlier maxima and minima.

[2] www.eecs.qmul.ac.uk/~simond/match/index.html
[3] www.sonicvisualiser.org

some preferences. Two participants reported that the acoustics of the medieval state were more satisfactory than that of the modern state to the interpretation of complex polyphonic music. One participant reported that singing in the modern state was less demanding than in the medieval state. In accordance with his previous singing experiences, this individual claimed to sing with less diligence in the modern state than in the medieval state, and that a longer reverberation time is more forgiving of small inaccuracies and defects in a performance, as heard from the audience. He also mentioned that as a listener, he would prefer the medieval state because it was "musically more satisfying". Despite their open design, two participants reported an unwanted "filtering" effect of the headphones , producing experimental conditions somewhat less comfortable than normal situations. Although open headphones let external sounds pass through, they still attenuate high frequencies, coloring the sound of one's voice, and the direct sound of the other singers.

6. CONCLUSION AND FURTHER WORK

In this study, all participants were able to correctly identify the more reverberant VAE indicating some perceptual validation of the auralization. However, there were some complaints about the usage of headphones which leaves room for improvement in future performance auralizations. There was no consensus as to which acoustic setting was optimal for the performance of music in the Ars Nova style.

Timbre, tempo, intonation, and vibrato did not seem to have been significantly affected by the acoustics. There is an indication, however, that the singers sang louder in the more reverberant modern configuration of the Great Chapel, but more data would be needed to strengthen this conclusion as it may be somewhat dependent on the style of the musical composition. Additionally, the impact of the acoustics on ensemble-specific features like inter-singer synchrony and intonation could be of interest in future analysis.

Finally, there are still recordings from this experiment which have not been analyzed, including those of a 4-voice ensemble interpreting a repertoire comprising one piece of monodic Gregorian chant and an additional choir configuration and position in the two acoustics. Analysis of these recordings could strengthen some of the preliminary findings of this paper as well as shed light on additional trends which may be a function of these additional variables.

ACKNOWLEDGEMENTS

Funding has been provided by UMR PRISM and Avignon Tourisme project IMAPI, the European Union's Joint Programming Initiative on Cultural Heritage project PHE (phe.pasthasears.eu), and the Paris Seine Graduate School Humanities, Creation, Heritage, Investissement d'Avenir ANR-17-EURE-0021 – Foundation for Cultural Heritage Sciences.

REFERENCES

[1] J.-M. Poisson, "Le palais des Papes d'Avignon : structures défensives et références symboliques," in *Les palais dans la ville*. Presses universitaires de Lyon, 2004, pp. 213–228.

[2] D. Tanay, "The transition from the ars antiqua to the ars nova: Evolution or revolution?" *Musica Disciplina*, vol. 46, pp. 79–104, 1992.

[3] K. Schiltz, "Church and chamber: the influence of acoustics on musical composition and performance," *Early Music*, pp. 64–80, Feb. 2003.

[4] P. Luizard, E. Brauer, and S. Weinzierl, "Singing in physical and virtual environments: how performers adapt to room acoustical conditions," in *Proc AES Conf: Immersive and Interactive Audio*, York, UK, Mar. 2019.

[5] T. Fischinger, K. Frieler, and J. Louhivuori, "Influence of virtual room acoustics on choir singing." *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 3, pp. 208–218, 2015.

[6] B. N. J. Postma and B. F. G. Katz, "Creation and calibration method of acoustical models for historic virtual reality auralizations," *Virtual Reality*, vol. 19, pp. 161–180, Sep. 2015.

[7] G. Colombe, "Au Palais des Papes : la chapelle Clémentine vue de l'intérieur," in *Mémoires de l'Académie de Vaucluse*. Forgotten Books, 1935, vol. XXXV, pp. 79–95.

[8] S. Mullins, V. Le Page, J. De Muynke, E. K. Canfield-Dafilou, F. Billiet, and B. F. G. Katz, "Preliminary report on the effect of room acoustics on choral performance in Notre-Dame and its pre-Gothic predecessor," *Meeting Acous. Soc. Am., J Acoust. Am.*, 2021.

[9] N. Eley, S. Mullins, P. Stitt, and B. F. G. Katz, "Virtual Notre-Dame: Preliminary results of real-time auralization with choir members," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2021.

[10] P. Iverson and C. L. Krumhansl, "Isolating the dynamic attributes of musical timbre," *J Acous Soc Am*, vol. 94, no. 5, pp. 2595–2603, Nov. 1993.

[11] M. Mauch and S. Dixon, "PYIN: A fundamental frequency estimator using probabilistic threshold distributions," in *IEEE Intl Conf on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 659–663.

[12] D. M. Howard, "Intonation Drift in A Capella Soprano, Alto, Tenor, Bass Quartet Singing With Key Modulation," *Journal of Voice*, vol. 21, no. 3, pp. 300–315, May 2007.