A computational approach to studying interdependence in string quartet performance

Panos Papiotis

TESI DOCTORAL UPF / ANY 2015

DIRECTORS DE LA TESI

Dr. Esteban Maestre
Dr. Xavier Serra Casals

Department of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain

Dissertation submitted to the Deptartment of Information and Communication Technologies of Universitat Pompeu Fabra in partial fulfillment of the requirements for the degree of

DOCTOR PER LA UNIVERSITAT POMPEU FABRA

This work has been supported by the EU FP7 FET-Open SIEMPRE (Project no. FP7-ICT-2009-C-250026), the Catalan Research Funding Agency AGAUR, as well as the EU H2020 project RAPID-MIX (H2020-ICT-2014-1 Project ID 644862).

To my father, for teaching me to love science.
To my mother, for showing me how it is done.

Acknowledgments

Of all the pages in this dissertation, these two are probably the ones that I had the hardest time writing, and the ones I had the least time to write them in.

When I began studying music I wanted to be a scientist, and when I began studying computer science I wanted to be a musician. Little did I know that there would be a time and place for me to combine the two. I am grateful to Xavier Serra for making the Music Technology Group the place it is today, and for giving me the opportunity to become a part of it and shape its future, surrounded by people with whom I share similar interests and motivations.

Even though this acknowledgments section is written in first-person singular, the reader will notice that the rest of this dissertation is written in first-person plural, and for good reason. The person that has been without a doubt the most influential in the definition, implementation, and conclusion of this work (besides myself!) is my principal supervisor Esteban Maestre. Together we spent endless hours debating and laying the foundation for the work presented in this dissertation. I am grateful and highly appreciative of his directness, honesty, and uncompromising dedication towards any task he is involved, as well as the effort and resources he has invested to make our research endeavors a reality. After all this work, I not only consider him my mentor but also my friend.

I am also grateful for having shared a large part of this journey with my Ph.D. colleague and friend Marco Marchini. He has proved to be an irreplaceable companion, from offering valuable advice and support to spending sleepless nights working on making the experimental recordings presented in this dissertation a reality. In the same vein, I am thankful to Alfonso Perez Carrillo for helping us shape the experimental procedure and organizing the experimental recordings, as well as Guillermo Navarro for providing valuable orientation on the topic of string quartet performance.

I would like to thank Prof. Marcelo Wanderley and Erika Donald for their support and help in organizing the experiments, as well as the CIRMMT and BRAMS labs in Montreal, Canada for hosting them. I am also grateful to the other McGill researchers who provided hands-on support in carrying out the experiments: Carolina Brum Medeiros, Marcello Giordano and Vincent Freour.

I am extremely thankful to Prof. Ralph Andrzejak for the many hours he dedicated advising and discussing with me on the applicability of interdependence estimation methods on music performance data as well as for putting me in touch with other researchers in the field of Nonlinear Time Series analysis so I could present and discuss this work with them. Similarly, I am thankful to Perfecto Herrera for his help and advice in carrying out the listening experiments presented in this dissertation.

I am also thankful to Oscar Mayor and Quim Llimona for the work they have put and continue to put into making RepoVizz what it is today; sharing office 55.312 with them has been an enjoyable and fruitful experience. Thanks are also due to Marina Alvarez Azon, for her help in cleaning data from motion capture recordings and preparing the dataset.

The last stretch of this Ph.D. was carried out in the great company of Sergi Jordà, Sebastian Mealla and Carles Julià. I thank them all, and especially Sergi, for giving me support and a new home within the MTG. Thanks are also due to Alba Rosado, Sonia Espi and Cristina Garrido for their support and help, as well as to Lydia Garcia and the rest of the DTIC administration.

There are many more people in the MTG with which I have shared many happy memories and fruitful discussions: Frederic Font, Alvaro Sarasua, Juanjo Bosch, Oriol Romani, Alastair Porter, Sergio Oramas, Giuseppe Bandiera, Marti Umbert, Andres Bucci, Marius Miron, Dmitry Bogdanov, Agustín Martorell, Mohamed Sordo, Justin Salamon, Jose Ricardo Zapata, Zacharias Vamvakousis, Sergio Giraldo, Stelios Togias, Sertan Şentürk, Sankalp Gulati, Gopala Krishna Koduri, Ajay Srinivasamurthy (also for helping out with latex math notation!), Ángel Faraldo, Daniel Gomez, Cárthach Ó Nuanáin, Martin Hermant, Mathieu Bosi. I thank them all, and apologize for the many others I have undoubtedly forgotten!

Throughout all these years I have been lucky to count with a number of people, both in Barcelona and in Greece, who I am proud to call my friends. They have provided immense moral support as well as understanding towards my often impenetrable “phd-ness”. I thank from the bottom of my heart my good friends Achilleas, Giorgos, Lefteris, Eva, Nikos, Ilias, Aggeliki, Stathis, Jason, Eliza (also for proof-reading this dissertation!), Vicky, Giannis, Giulia, Joan, Nuria, Ayse Naz, Alberto, Lauren, Ricardo, Pascal, Sytse, Silvia, Dimitra, Alex and all of the others who have supported me, played music with me, made me laugh and think in the past 5 years and for many years to come.

For the last few people in these acknowledgments, I have no proper way of thanking enough, but here goes: my parents Kostas and Kalliopi, for being my heroes, my reference at every happy or difficult moment in my life, and the providers of infinite, unquestioning love and support. My sister Hara, the most talented person I know, for always understanding me and having me in her thoughts.

Finally, my partner and soon-to-be wife Elena, for making these years the best and most fulfilling years of my life.

Abstract

This dissertation proposes a computational data-driven methodology to measure music ensemble interdependence - the degree to which musicians interact and influence each other's actions in order to achieve a shared goal - using a string quartet ensemble as a case study.

We present the outcomes of an experiment on music ensemble interdependence, where we recorded the members of a professional string quartet performing exercises and excerpts of musical pieces in two conditions: *solo*, where each musician performs their part alone, and *ensemble*, where the entire quartet performs together following a short rehearsal. During the performance we acquire multimodal data in the form of audio recordings, motion capture data from sound-producing movements and upper body movement, as well as high quality video. All of the recorded data have been published online as an open research dataset.

From the acquired data, we extract numerical features in the form of time series that describe performance in terms of four distinct musical dimensions: intonation, dynamics, timbre, and timing. We apply four different interdependence estimation methods based on time series analysis - Pearson correlation, Mutual Information, Granger causality and Nonlinear Coupling coefficient - to the extracted features in order to assess the overall level of interdependence between the four musicians for each performance dimension individually. We then carry out a statistical comparison of interdependence estimated for the ensemble and solo conditions. Our results show that it is possible to correctly discriminate between the two experimental conditions for each of the studied performance dimensions. By computing the difference in estimated interdependence between the ensemble and solo condition for a given performance dimension, we are also able to compare across different recordings in terms of the established interdependence and relate the results to the underlying goal of the exercise.

We additionally study the aural perception of music ensemble interdependence, assessing the capability of listeners to distinguish between audio recordings of ensemble performances and artificially synchronized solo performances as a function of the listeners' own background and the performance dimension that each recording focused on.

The proposed methodology and obtained results explore a novel direction for research on music ensemble interdependence that goes beyond temporal synchronization and towards a broader understanding of joint action in music performance, while the shared dataset provides a valuable resource that serves as a foundation for future studies to build upon.

Resumen

Esta disertación propone una metodología computacional basada en datos para medir la interdependencia entre los intérpretes de un conjunto musical - el grado en que los musicos interactuán e influyen sus acciones para conseguir un objetivo común - utilizando un cuarteto de cuerda como caso de estudio.

Presentamos los resultados de un experimento sobre interdependencia en un conjunto musical, donde registramos a los miembros de un cuarteto de cuerda profesional interpretando ejercicios y fragmentos de piezas musicales en dos condiciones, *solo*, en la que cada músico interpreta su parte sin acompañamiento, y *ensemble*, en la que todos los miembros del cuarteto tocan juntos despues de un ensayo. Durante la interpretación, grabamos datos multimodales en forma de audio, captura de movimientos productores de sonido y de la parte superior del cuerpo, así como vídeo de alta calidad. Todos los datos adquiridos han sido publicados online en forma de una base de datos de acceso abierto para fines de investigación.

De los datos adquiridos, extraemos descriptores numéricos en forma de señales temporales que describen la ejecución en relación a cuatro dimensiones de la interpretación musicales: entonación, dinámica, timbre y tempo. Aplicamos cuatro métodos diferentes de estimación de interdependencia basados en análisis de series temporales - Pearson correlation, Mutual Information, Granger causality y Nonlinear Coupling coefficient - sobre los descriptores obtenidos para estimar el nivel total de interdependencia entre los cuatro músicos para cada una de las dimensiones de interpretación. A continuación, llevamos a cabo una comparación estadística de la interdependencia estimada para las condiciones de conjunto y solo. Nuestros resultados muestran que es posible discriminar correctamente entre las dos condiciones del experimento para cada una de las dimensiones de interpretación estudiadas. Mediante el cálculo de la diferencia en la interdependencia estimada entre las condiciones de ensemble y solo para una dimensión de interpretación, también podemos comparar entre diferentes grabaciones en términos de la interdependencia establecida y relacionar los resultados con el objetivo fundamental del ejercicio.

Además, estudiamos la percepción auditiva de la interdependencia en el conjunto musical, evaluando la capacidad de los sujetos de distinguir entre grabaciones de audio de interpretaciones ensemble y otras creadas artificialmente sincronizando interpretaciones en solo en función de los conocimientos del sujeto y de la dimensión de interpretación en que cada grabación se centraba.

La metodología propuesta y los resultados obtenidos exploran una nueva dirección para la investigación de conjuntos musicales que va más allá de la sincronización

y hacia una comprensión más amplia de la acción conjunta en la interpretación musical, mientras que la base de datos compartida proporciona un valioso recurso que sirve de fundamento sobre el que se construyan nuevos estudios.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1. Music performance as a research discipline

Music is a social experience[1]. Composers share their work with performers, which then interpret it and present it to an audience. Although it is not uncommon for the same person to fulfill all of the roles involved (composer, performer, audience), in most cases music implies a certain degree of interaction between humans (which they may or may not be aware of, and may or may not be voluntary).

From this viewpoint (as seen in Figure 1.1), we can highlight several different interactions that are present within music; such as composer-performer interaction, performer-audience, composer-audience, et cetera. Since each role can be filled by more than one person (i.e. multiple composers/performers/listeners), the spectrum widens when inter-role interactions are considered.

These interactions - and the tasks around which they revolve - are highly appealing topics for research, as they can be studied to reveal valuable information on human cognition, creativity, and communication (for reference, see the parallels drawn between music and language as research topics Patel, 2010). Of course, the meaning of the word "interaction" greatly varies with the underlying context: a composer can *interact* with the audience by conveying an emotion through the composition; while a maestro can *interact* with the tympanist of an orchestra via the movement of his baton in order to induce an acceleration to the performed

[1] *"Music is..."* must be one of the most common opening statements for doctoral dissertations in our field. Of course, it is nigh impossible to define what music is with one sentence, and doubly so when that one sentence has to represent your own body of work as well. It is important to clarify that this particular opening statement serves as an orientation rather than a definition.

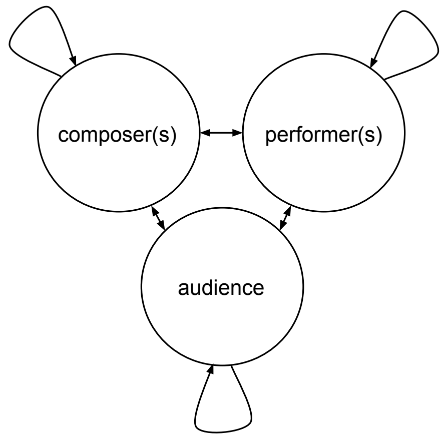


Figure 1.1: Diagram of social interactions in music involving the composer(s), performer(s) and audience, both between groups and within groups.

tempo. Clearly, these two *interaction phenomena* are fundamentally different and most likely require different research methodologies in order to be studied.

Among the various phenomena that are of scientific interest within musical interaction, one can argue that **performance** appears as the central connector; a composed piece needs to be *performed* in order to reach an audience (even if the performance consists of a series of instructions carried out by a computer). Of course, even by placing the focus on musical performance alone, the breadth of topics contained within is still immense when considering different music genres, instruments, and the several stages of preparation that precede the actual performance of a piece in front of an audience.

What is also particularly interesting in treating music performance as a subject for scientific research, is that the choice of discipline with which to approach a topic is not always straightforward. As an example, let us select a single case as a hypothetical topic for scientific study, such as a pianist performing a formal concert of a western classical period piece, along with a specific research problem: is there a direct relationship between a listener's emotional response to a performance, and the expressive intentions of the performer? Now consider the different disciplines from which the problem can be approached: a musicologist might analyze features of the performed composition (tempo, mode, dynamics, etc) and the performer's deviations from them in conjunction with the listeners' features (such as development, musical training, context). A psychologist might carry out a qualitative study based on interviews and questionnaires, and interpret them as parameters of an existing theoretical model. A neuroscientist might perform quantitative measures of physiological signals in both the performer and the listeners, and investigate similarities in their affective states based on physiological response.

More often than not, researchers from all these disciplines will join their forces to approach the topic holistically; and since a significant amount of technical knowledge is required for several of the above tasks (usage of sensors and measurement devices, computational tools, statistical analysis of results) it is almost given that engineers will join the research team. Besides the methodology used, it is not uncommon to see findings originating from one field used or challenged in a completely different field. Therefore, music performance research does not simply invite multidisciplinarity, but rather requires it.

The mixture of different disciplines involved in music performance research also carries implications on the type of research that is carried out. **Top-down** approaches start from a hypothesis informed from a theory, and then perform experiments under controlled conditions in order to collect observations that either reject or validate the hypothesis. On the other hand, **bottom-up** approaches have a more exploratory character, where observations collected directly in the phenomenon's natural environment (also called *ecological* conditions) lead to the discovery of patterns, which culminate in the proposal of some tentative hypotheses as components of a theory. Similarly to the choice of discipline, music performance research often follows a **mixed** approach throughout the progression of a project: experiments carried under controlled conditions yield observations that have not been anticipated, which in their turn develop into new theories through studies carried out in ecological environments.

1.2. Ensemble music performance

Ensemble performance refers to a number of musicians performing together as a group, the size of which can vary from a simple duet to any number of musicians, such as a complete symphonic orchestra. Larger ensembles are usually led by a conductor who acts as a leader in terms of temporal coordination, evaluates the performance and provides additional instructions on how it should be carried out. On the other hand, small ensembles are typically conductor-less, and such tasks are placed upon the musicians themselves. In small ensembles, each performer usually has their individual part, while in large ensembles it is common for given sub-sections of the same instrument family to have identical parts - an example of this are first and second violin sections within a symphonic orchestra.

Ensemble performance and the skills associated with it encompass an undeniably significant subset of music performance overall, due to its social nature. Music education is firmly rooted in collective performance scenarios such as singing or playing in groups, and in fact participation in ensembles such as chamber music groups or student orchestras are often considered to be the first step towards a professional career in music. Outside formal scenarios, one would be hard-pressed to find amateur musicians that are either unable or unwilling to perform together with others.

Despite this, the amount of published research focusing on ensemble music performance is small compared to research published on individual performance. While there is a gradually growing body of work dedicated solely to ensembles, it could be said that the volume of research is spread somewhat thinly over the large surface of sub-topics within. Chapter 2 of this dissertation contains a review of the existing literature, which spans different types of ensembles, aspects of the performance that the study focuses on, as well as involved academic disciplines.

Given how ensemble performance lies at the intersection of music performance and social interaction, research outcomes are not necessarily limited to music. Neighboring topics such as language, as well as performing arts such as theater and dance share many common characteristics; and thus findings originating from research on one field can inform or influence theoretical approaches in another, while quantitative approaches can be adapted and applied as tools for scientific experimentation. An example of both of these cases can be found the up-and-coming field of Social Signal Processing (Pantic et al., 2011), where multimodal data originating from human non-verbal behavior are automatically analyzed using signal processing techniques to study social interaction and derive context-aware causal relationships between interacting agents.

1.3. Synchronization and Interdependence in ensemble music performance

Naturally, the most fundamental concept of ensemble performance is **synchronization**, maintaining the same tempo across all members of the ensemble. This is achieved via a process which the music perception field describes as *sensorimotor synchronization* (Repp, 2005), the rhythmic coordination of action (i.e. sound-producing movements) with tempo perception. Sensorimotor synchronization, in its most basic form, describes two cognitive mechanisms that account for short-time and long-term adjustments to an external rhythm: *phase correction* and *period correction*; through studying the onset times of tapping performed while listening to an external rhythm, research on Sensorimotor Synchronization tries to formulate mathematical models to describe these two mechanisms.

In the case of a solo performer, tempo perception might be solely affected by the performer's internal timekeeping, or by an unchanging external influence such as a metronome device or a pre-recorded performance. In ensemble performance, the complexity of sensorimotor synchronization increases considerably since both tempo perception and rhythmic coordination of action exist in a closed feedback loop that involves every member of the ensemble.

Synchronization is but the first requirement towards achieving joint musical expression. In this dissertation we use the term **interdependence**, loosely borrowed from the field of time series analysis, to refer to a broader phenomenon in ensemble music performance: the *coordinated variation of expressive aspects of the performance by the members of a musical ensemble*. From an expressive point of view (in contrast to a purely executive one), musicians performing together in an ensemble must not only "keep time" but also jointly shape the more nuanced aspects of their performance in order to achieve a (presumably) shared aesthetic goal (Keller, 2008). These "more nuanced" aspects of the performance clearly are defined by the expressive affordances of the instrument; while a piano performer can mainly modulate dynamics and the articulation of musical phrases (via variations in timing and keystrokes), a singer can additionally modulate both intonation as well as timbre simultaneously and in a continuous manner. Interdependence, or the coordinated variation in these distinct dimensions of the performance by the musicians, is what eventually results in joint musical expression.

There is no single accepted set of terms to refer to these distinct, co-existent qualities of a musical performance; while the terms intonation, dynamics, timbre and timing are ubiquitous in the music performance literature, they are often presented together with more abstract, compound concepts such as *articulation* or *phrasing* (both of which may combine dynamics, timbre, and timing). In this dissertation, we use the term **performance dimensions** to refer to four basic, simultaneously present aspects of a musical performance: *intonation*, *dynamics*, *timbre* and *timing*.

There are several factors that can contribute to, or hinder interdependence in music performance: the technical difficulty of the piece in relation to the capabilities of the performers, the conditions under which the performance takes place, the performers' familiarity with the piece and with each other, and their state during the performance, to name a few. Besides practical factors, joint musical expression assumes that a common goal is shared by the musicians - an assumption that cannot be always considered fully true in the performance of every ensemble. We can therefore point out a key difference between the concepts of synchronization and interdependence: the former refers to the basic skill of *playing in time with* *each other* required for ensemble performance, while the latter requires a different type of musical expertise, the ability to *communicate and perceive* the expressive qualities of musical performance.

1.4. String quartet performance as a case study

This dissertation directs largely all of its methodological focus on a particular type of music ensemble called the *string quartet*, a small conductor-less ensemble of four musicians (typically two violins, a viola and a cello) where every musician has their own individual part. It is one of the most prominent chamber ensembles in classical music, with most major composers from the mid to late 18th century onwards, writing pieces for string quartets. String quartets do not have a *de facto* leader: although the first violinist is traditionally considered the 'leader' of a string quartet, this cannot be always expected to be the case, and there is no guarantee that he/she will have the greatest decision-making influence in the group (Young and Colman, 1979).

There are several characteristics of string quartets that constitute them fitting as subjects for empirical research on ensemble music performance. To begin with, the majority of both their educational as well as concert repertoire is firmly rooted on western classical music, meaning that the musicians are expected to follow the underlying score quite closely. This makes it possible to objectively analyze their performance, since their actions can be contrasted to a common reference. Thus, performances carried out under different experimental conditions can be compared with each other.

The size of the ensemble also holds a significant advantage: it is not too large as to render analyses of synchronization and interdependence infeasibly complex (such as the case of an orchestra), while still being more complex than the simplest ensemble possible (the duet). The lack of a conductor or a *de facto* leader also means that, ideally, all members potentially share equal responsibility in achieving the ensemble's common artistic goal.

Finally, the expressive affordances of bowed string instruments make for a rich set of parameters that can be adjusted to shape the performance: sound is generated via continuous excitation of the strings, which makes dynamics and timbre continuously evolving aspects of the performance rather than the result of a discrete event (such as plucking or striking). Furthermore, bowed string instruments are fretless and are therefore capable of producing continuous pitch as well; and the harmonic consonance of the overall sound is achieved through constant adjustments to each musician's intonation through the use of different *tuning temperaments* (fine adjustments to the note intervals according to the harmonic context).

1.5. Scope of this work

The goal of this dissertation is to measure and study the interdependence among the members of a string quartet through computational means. In this section we will narrow down the scope, first by defining the questions we set out to answer, and then by outlining the methodology we designed in order to do so and the additional outcomes that followed.

1.5.1. Problem statement

Research on ensemble performance is typically centered around the two core concepts previously introduced in Section 1.3, *sensorimotor synchronization* as a prerequisite for joint action, and *interdependence* as a prerequisite for joint musical expression. While sensorimotor synchronization is rooted in complex motor processes and cognitive human functions that are still not fully understood, the phenomenon itself is relatively straightforward to observe and quantify by obtaining the onset times of each performed note and computing the asynchrony between performers. Interdependence is more difficult to observe and quantify, as it involves additional dimensions of the performance beyond timing -intonation, timbre and dynamics- which to this day lack specific low-level definitions.

This highlights a methodological imbalance in analyzing ensemble performance: while sensorimotor synchronization studies are built upon a mathematical framework informed by quantitative tools both for descriptive as well as theoretical modeling purposes, studies on joint musical expression and interdependence are mainly limited to expressive timing since it can be quantified in a straightforward way. Due to the lack of a reliable methodology for **observing** and **measuring** interdependence in a musical ensemble beyond synchronization, the literature mostly focuses on the question of *how* musicians interact, studying the underlying cognitive mechanisms that the performers employ to bring about the optimal co-variation of expressive performance parameters. In doing so, it is often assumed that expressive interdependence has indeed been achieved by the ensemble in the particular recording that is being examined - an assumption that cannot be always considered fully true in every ensemble's performance.

We would therefore like to pose a different question, namely *whether* musicians are interacting:

> **Given a certain performance by an ensemble, is it possible to measure *whether* the musicians are indeed interacting and adapting to each other to achieve a coordinated result, in terms of distinct dimensions of the performance (intonation, dynamics, timbre, timing)?**

Providing an answer to this question requires a methodology for **measuring** interdependence in the musicians' expressive choices: this methodology is the core of this dissertation.

1.5.2. Methodology

The essence of our methodology can be summarized as follows. We capture the performance of a string quartet in two discrete conditions: *solo*, where each member performs on their own, and *ensemble*, where the quartet performs together as an ensemble. We then employ computational tools to compare the two conditions and measure the difference between them in terms of interdependence.

The realization of this methodology can be further broken down in three steps. First, the performance of the quartet is captured using different data modalities simultaneously, with the main modalities being *sound* and *movement*. Second, raw data recorded during the performance are computationally analyzed to obtain numerical descriptors that approximate a musical representation of the performance. Third, time series analysis methods capable of estimating interdependence are applied to these numerical descriptors to assess musical interdependence between the members of the ensemble. Figure 1.2 presents an overview of our methodology, starting from the low level of abstraction in the data acquired in performance capturing, moving towards mid-level representations of the acquired data during performance description, and concluding with high-level, abstract numerical representations of interdependence in the last step.

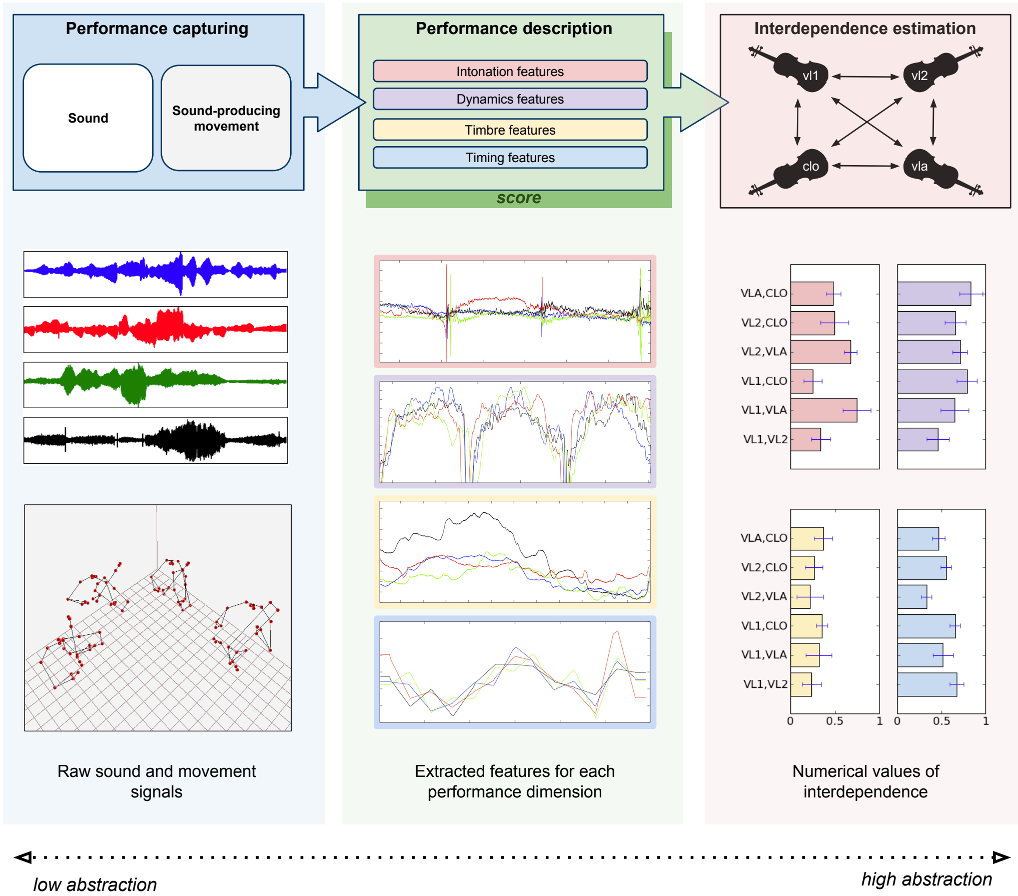


Figure 1.2: Overview of the ensemble interdependence analysis methodology presented in this dissertation, demonstrating the different levels of data abstraction in each step of the methodology.

Performance capturing. To devise and test our proposed methodology, it was necessary to work with data from real performances captured under experimental conditions. Two main considerations were taken into account regarding the experimental design of the capturing conditions. The first one was ensuring that each recording has a simple shared goal that is clearly understood by the performers. The second consideration was to capture data that would reflect the studied phenomenon in the most unambiguous manner possible, focusing around having a clear condition where *no* interdependence among the musicians exists, and a case where *some* interdependence exists. For this reason we also capture the performance of each musician performing individually, without having access to, or prior knowledge of the complete ensemble's score.

The actual data captured during the performance belong to two categories: the *produced sound* of the ensemble (recorded using individual pickup microphones fitted on the instruments' bridge and ambient microphones), as well as the *sound-producing movement* of the musicians during the performance (recorded using motion capture technology). By capturing the sound-producing movement, or *instrumental gestures* of the performers, we gain additional information on complex aspects of the produced sound that is difficult to obtain solely from audio analysis. Examples of this are the distance of the bow-string contact point from the bridge which strongly affects the *timbre* dimension of the produced sound, or the applied bow force and velocity which provide valuable information in determining the exact timing of note onsets and offsets.

Performance description. In the second step of our methodology, the performance data are processed and analyzed in order to describe the performance of the string quartet in terms of the four performance dimensions (intonation, dynamics, timbre, and timing). In order to do so, we first align the underlying musical score to each performance, obtaining precise note onset and offset times in the recordings for each performed note. This provides a common reference through which different performances can be compared to each other.

Following score alignment, the audio and movement information is analyzed in order to extract time series features capable of describing musical performance. These descriptor features (such as *pitch*, *sound intensity*, *bow-bridge distance*, *spectral crest*) are combined with information from the aligned score (such as bar positions, initial tempo, expected pitch of the performed note, etc) to produce the final set of performance descriptors.

Interdependence estimation. Finally, the last step of our methodology deals with estimating the amount of interdependence that is established during the ensemble's performance. Since we are essentially dealing with time series, in this step we apply a set of existing interdependence estimation methods fit for time series features (*linear correlation*, *mutual information*, *granger causality*, as well as *nonlinear coupling coefficient*).

Given the fact that the time series features for each performance dimension describe the same phenomenon across all four musicians (i.e. pitch deviation, sound intensity, etc), they are bound to share common characteristics (e.g. linearity, stationarity, autocorrelation, et cetera). Moreover, since the performances originate from the same score, salient events (such as note start/end times) often occur simultaneously in the performances of different musicians. For these reasons, every estimation of interdependence that we carry out is expected to yield non-zero values.

In order to have an idea of what is the numerical equivalent of "zero interdependence" for each recording, we utilize the recordings in the *solo* condition as a baseline for which interdependence is estimated identically to the ensemble recordings; the difference between *solo* and *ensemble* interdependence is what we use as our final estimation of interdependence. We additionally carry out a set of statistical tests in order to assess whether the difference between the solo and ensemble conditions is significant.

1.5.3. Additional work

Beyond the design and evaluation of the core methodology, there was a number of additional outcomes and paths that were explored as part of this dissertation. The task of designing a methodology to observe/measure a phenomenon is bound to lead to findings, the first of which is -strangely enough- evidence that the phenomenon exists in the first place. This was the case for our work as well; it is reasonable to assume that e.g. string quartet musicians influence each other's intonation in general, but it is not easy to assess whether this is indeed happening in a particular recording that is being analyzed. Moreover, by studying distinct dimensions of the performance, it can also prove possible to observe how interdependence in one dimension (e.g. intonation) varies for different scores.

Naturally, the motivation behind observing these phenomena is fueled by the inherent knowledge that humans are not only indeed capable of joint expression, but also capable of perceiving and appreciating it as listeners; if not, how could the performance of musical ensembles be such a rich field both for art as well as science? Therefore, besides designing a methodology to observe/measure musical interdependence, we also dedicated part of this work to explore the listeners' ability to perceive it as well.

Finally, we consider one of the most concrete outcomes of this work to be the creation and sharing of a large research dataset on ensemble music performance, detailed in Chapter 3. One of the obstacles to overcome during our initial research efforts was the lack of reliable data on which we could apply our methods. Carrying out large scale experiments with multiple musicians where multimodal data are acquired is a costly, laborious process not only during- but also post-acquisition, as this large amount of data must be processed, cut, synchronized, manually "cleaned" (in the case of motion capture data) and annotated. During this entire process, one of our clear goals was to contribute to the field of ensemble music performance by sharing this large dataset, and prepare it in such a way that it can prove useful to other fields of research as well.

1.6. Thesis timeline

As a starting point, we carried a pilot study on violin duets of both professional as well as amateur performers, recording part of the duet's existing repertoire in two simple conditions: *solo*, where each musician performed alone, and *ensemble*, where the duet performed together. Outlined in (Papiotis et al., 2012b, 2011), these pilot experiments were instrumental to the future progress of this work, implementing the first iteration of our technical setup (which we will revisit later in this section) but perhaps more importantly revealing the limitations in our initial experimental procedure.

In her literature review of music performance research (Palmer, 1997), Caroline Palmer sums up two important challenges that are almost always present in the analysis of experimental data, and that were certainly present in our pilot recordings. First, it is difficult to select performances that are consistently representative of the phenomenon that is being studied:

> *"A methodological problem is determining which performances should be considered representative, given the large variations that can occur among competent performances of the same music. [...] A similar representativeness problem arises in choice of musical stimuli. Because of complexity issues, experimenters often use simplified or reduced musical compositions, [...] through which evidence is converged from both small and large sample studies conducted with different musical stimuli."*

What the above quote illustrates, is that simply having the musicians perform pieces from their existing repertoire may not be fully representative of the phenomenon we wanted to study. Instead, we considered the use of reduced compositions, whose goal was to leave as few doubts as possible as to whether they can be considered representative examples of the studied phenomenon.

A second challenge, related to the above statement as well, had to do with separating meaningful instances of recorded data from "noise":

> *"The wealth of data from a single performance [...] results in problems of separating signal from noise. Current computer music technology relies heavily on movement-based information and records only event onsets, offsets, and their relative intensities from electronic- or computer-monitored musical instruments. Despite the reduction of information, problems with separating the signal -performance*

expression [2]*- from random noise fluctuations remain.*

This brings out a couple of important points in relation to our problem: first, that the use of multimodal data sources gives us the luxury of choosing the most informative (or least "noisy") data source in order to have a clear picture of the musicians' actions, while computational performance description techniques can help in reducing redundant information; and second, that in order to compare performances and extract meaningful expressive parameters (or as Palmer calls them "Performance expression"), a steady reference combined with repetitions of the same composition is required.

In the search of *meaningful* data that are *representative* of joint musical expression, we carried out a new round of experimental design and implementation. This process posed two challenges:

- **Experimental design and materials**, where representative compositions capable of producing meaningful data had to be chosen or produced, and
- **Technological framework implementation,** where our technical setup had to accommodate for the simultaneous acquisition of data originating from multiple modalities, from four musicians simultaneously, while maintaining relatively ecological (i.e. non-disruptive) conditions for the performers.

In order to carry out the **experimental design**, we invited a professional string quartet performer (Navarro, 2011) to discuss the process of establishing interdependence in a string quartet. Through an intensive workshop, he explained how the ability to collectively perform expressively as a single unit should not be taken for granted; and that string quartets musicians, regardless of how skilled they are individually[3], always have to dedicate special efforts on strengthening the interdependence among them, for different dimensions of the performance. This process often involved working on very subtle aspects of the performance - and for performance dimensions such as intonation or timbre, the difference might not be even perceived by an 'untrained' ear.

Along with his knowledge and guidance, the invited musician also provided us with a document invaluable for this work: an exercise handbook for string quartets (Heimann, 1958). This handbook consisted of short exercises and piece excerpts

[2]Quoting from the same text: *"Performance expression refers to the large and small variations in timing, intensity or dynamics, timbre, and pitch that form the microstructure of a performance and differentiate it from another performance of the same music."*

[3]In an interesting anecdote, the invited musician shared a story with us of going to see a string quartet concert that was composed by four extremely renowned musicians, each one individually considered a master of their instrument. However, he was very disappointed in the concert, as the musicians were merely playing their parts together in time but not really *listening* to each other!

divided in categories, each focusing on a specific dimension of the performance, with the goal of strengthening the interdependence among the performers of a string quartet. This very document formed the basis of our experimental materials. The entire experimental design process, including a description of the materials, is detailed in Chapter 3.2.

In order to implement the **technological framework** for multimodal data acquisition, we iterated over several components of a setup already designed in (Maestre, 2009; Perez Carrillo, 2009). This setup involved the simultaneous acquisition of data in multiple modalities. We acquired individual audio from each musician via a bridge pickup microphone, while simultaneously recording the entire quartet by means of a large diaphragm cardioid microphone as well as a binaural dummy head. Motion information coming from the musicians' upper bodies as well as detailed instrumental gesture data from the instruments' bodies and bows was acquired using two motion capture systems simultaneously. Finally, experimental sessions were video recorded as a reference, and qualitative data through the use of questionnaires was acquired after select recording sessions.

Besides data acquisition, our technological framework included a series of further processing steps: cleaning the acquired data, synchronizing data streams recorded with different devices, as well as annotating and aligning each recording with its underlying musical scores. Additionally, the data capturing and preparation process highlighted the need for equally important multimodal data storage, visualization, and sharing components. Our efforts on this field, culminating in the release of a large multimodal dataset of string quartet performance, are detailed in Chapter 3.3.

This work was carried out as a collaborative project. Several researchers both at the Music Technology Group of Universitat Pompeu Fabra, as well as researchers and lab assistants at the Centre for Interdisciplinary Research in Music Media and Technology of McGill University dedicated a significant effort to making this work possible; references to their work are present throughout this document, as well as in the Acknowledgments section.

1.7. Thesis outline

The organization of this dissertation is as follows. In Chapter 2 we introduce and review existing work on three research topics that are relevant to this thesis: *Music Performance*, placing additional focus on musical ensembles; *Expressive performance*, which deals with the quantification of musical expression using computational means; and *Interdependence estimation*, which outlines the main computational methods for quantifying interdependence among time series features.

In Chapter 3, we outline the experimental design procedure. We start by re-visiting the preliminary work carried out in a pilot study. We then explain the design of the conditions under which our main experiment was carried out, discuss in depth the selection of musical materials (scores) highlighting their relations to important musicological concepts involved, and present the technological framework for data acquisition and processing, from the stage of preparation until the release of an openly available dataset of the corpus acquired for the purposes of our subsequent analyses. We outline our implementation of a score-performance alignment algorithm that was used to obtain accurate timing information for note onset and offset times in each recording. Finally, we present the dataset that was released as part of our research efforts, along with the online platform that was implemented in order to share it.

In Chapter 4 we present our methodology for quantifying interdependence among the musicians of a string quartet, for four distinct dimensions of the performance (intonation, dynamics, timbre, and tempo). We discuss the feature extraction algorithms that were applied to describe each dimension of the performance (both in terms of audio as well as motion), and present the extracted features as well as the interdependence methods that we applied to the extracted features. Finally, we provide a detailed analysis of the obtained results, and discuss their implications. We additionally present the results of a listening experiment that investigates the perception of musical interdependence by listeners, and the factors that affect the listeners' capability to perceive interdependence.

Finally, in Chapter 5 we offer our conclusions. We summarize the contributions of this dissertation in terms of experimental design, multimodal data acquisition, the shared dataset, our computational methodology, as well as our perceptual studies. We discuss the limitations of our work, and propose directions for future research.

Chapter 2

BACKGROUND

In this chapter we introduce the scientific background for three research areas relevant to this dissertation. First, we present the state of the art on the *Music Performance* field, focusing on the work that is most relevant to this dissertation both in terms of methodology and subject; special attention is placed on the relatively recent studies that specifically target *Ensemble Music Performance*.

We then present the state of the art on the field of *Expressive music performance* and more specifically the performance description methodologies within, as they are strongly related to the computational methodology followed in this thesis. In the performance description subsection we make specific mention to the field of *Music Information Retrieval*, as a large part of the signal processing and feature extraction techniques applied are common to both fields.

Finally, we briefly introduce and describe the state of the art on *Interdependence estimation* for time series, focusing on the most ubiquitous methods and their characteristics.

2.1. Music performance

2.1.1. General music performance literature

Music performance, as any other form of performance art, has traditionally been the subject of systematic study. Published treatises on proper performance practice have been circulating as early as the 18th century (Mattheson, 1739; Quantz, 1752; Bach, 1753; Mozart, 1756) and remain relevant to this day. However, according to

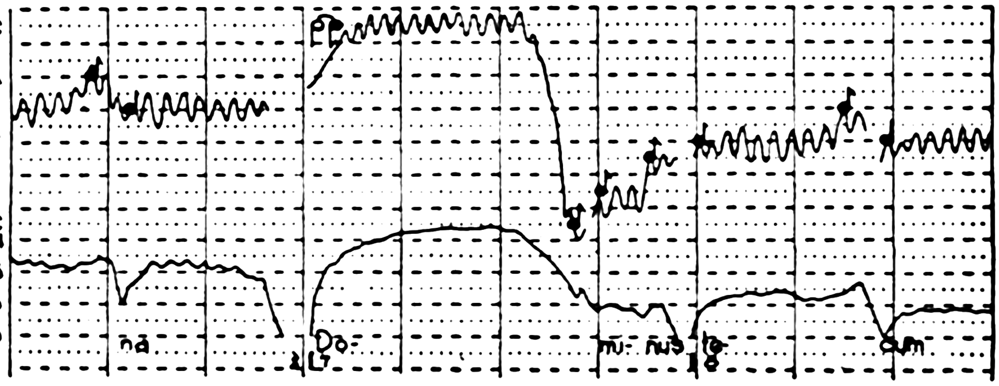


Figure 2.1: An example of pitch (top) and sound intensity (bottom) data from Seashore's research at the beginning of the 20th century; this segment originates from a vocal performance of Bach-Gounod's *Ave Maria*.

Madrid (2009), the study of music performance had "a narrow meaning within the musicological context: the examination of how the music score as a text could be realized into sound, which eventually raised questions of correctness in interpretation, historical authenticity, and the role of oral transmission in reproducing the *true* spirit of said music text".

It was not until the early 1900's and the advent of the then-new academic discipline of Psychology that research on music performance grew out of the confines of musicology and entered traditional scientific circles. Utilizing empirical research methodologies, initial efforts were concentrated on analytical measurements of musical performance, carried out by a research group headed by Carl Emil Seashore at the University of Iowa (Seashore, 1936). Through the use of a novel camera-based system that captures hammer and foot pedal movement, as well as existing devices such as Olaus Henrici's *Harmonic Analyzer* (Tinker, 2006) for the case of violin and voice, Seashore and his team were able to carry out precise measurements of dynamics, timing, and intonation information from real performances. Figure 2.1 shows an example of the data Seashore based his analyses on, for a vocal performance (from Seashore, 1938).

Consistent with its beginnings, music performance research has traditionally been dominated by studies around the measurement of performance. In an exhaustive state of the art review by Gabrielsson (1999), approximately 220 out of 500 papers were found to be based on performance measurements. Gabrielsson additionally reports that many of the papers dealing with motor processes and models of performance also referred to measurements of various kinds. Of course, performance measurements are not an end unto themselves, but rather examples of empirical research, where findings from direct and indirect observation of music performance

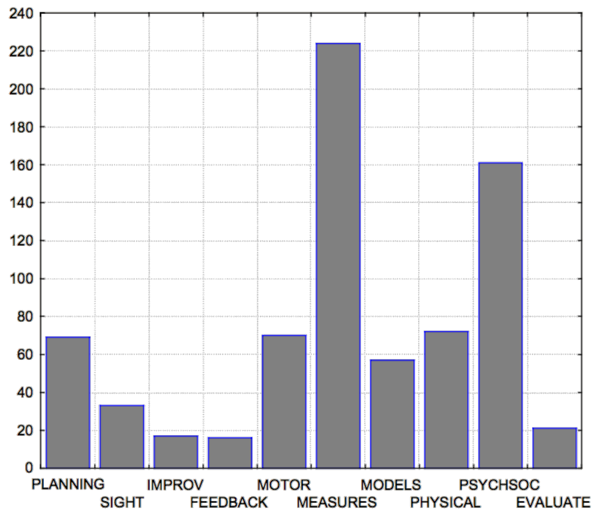


Figure 2.2: Number of papers within different areas of music performance research (from Gabrielsson, 1999): SIGHT = sight reading; IMPROV = improvisation; MOTOR = motor processes; MEASURES = measurement of performance; PHYSICAL = physical factors; PSYCHSOC = psychological and social factors; EVALUATE = evaluation of performance.

are used to discover patterns, formulate hypotheses and build theoretical models.

Gabrielsson structures the Music Performance literature topics as follows: *performance planning*, which deals with acquiring an adequate mental representation of the piece to be performed as well as practicing; *sight reading* (also known as *prima vista*), which deals with performing from a score without previous knowledge or practice of the piece; *improvisation*, the meaning of which is simultaneously self-explanatory and regularly challenged (see Pressing, 1988); *feedback*, which deals with the different types of feedback received during performance (auditory, visual, proprioceptive); *motor processes*, which deal with the cognitive skills as well as physical capabilities that control and influence movement in music performance; *measurements of performance*, which we already presented; *models of music performance*, which refer to theoretical as well as computationally derived rules that explain and predict expressive deviations in music performance; *physical factors*, which deal with medical effects of music performance on the performer's body and hearing; *psychological and social factors* that deal with musical development, personality, occupational effects, as well as performance anxiety; and finally *evaluation* of performance in different environments (by teachers, critics, listeners, and other musicians). It is important to note that, with very few exceptions, published research studies deal with western tonal music, mostly art ("classical") music. The volume of published articles in each of these topics in Gabrielsson's 1999 literature review article can be seen in Figure 2.2.

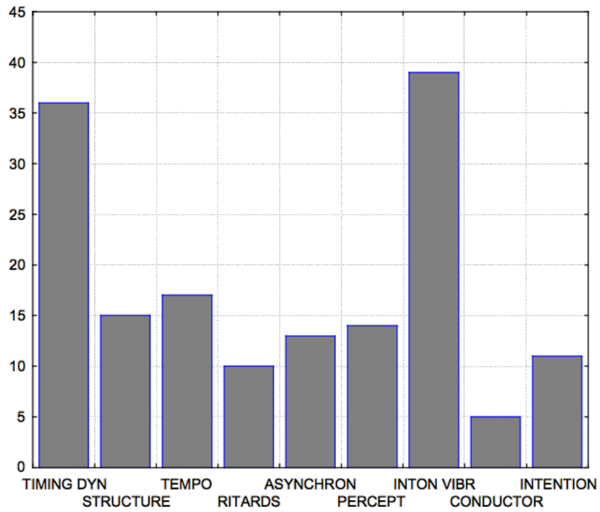


Figure 2.3: Number of papers on the topics within the *measurement of performance* category (from Gabrielsson, 1999): TIMING DYN = timing and dynamics; ASYNCHRON = asychronization; PERCEPT = perceptual effects; INTON VIBR = intonation and vibrato; INTENTION = intention and performance.

This organization of topics is, for the most part, consistent with other literature reviews on the field of music performance. Differences usually arise from the background of the reviewer and the purpose of the review. An article by Palmer (1997) focuses on cognitive aspects of music performance such as memory retrieval, anticipatory planning, and motor control; Rink (2002, 2003) provides a musicologist's point of view on the literature, while Parncutt and McPherson (2002) attempt to bridge the gap between music performance research and music practice itself. Finally, Widmer et al. (2003), Goebl et al. (2008) and De Poli (2004) place a strong focus yet again on performance measurement, this time in the context of computational analysis of expressive music performance for the purposes of modeling; we will delve deeper into the subject in the upcoming Section 2.2.

Before moving on to research on the performance of musical ensembles, we will present the state of the art on two types of empirical music performance studies that are of special importance here: (i) studies that measure performance in terms of the four dimensions analyzed in this dissertation (intonation, dynamics, timing, timbre) and (ii) studies that measure the performer's movement.

2.1.1.1. Measurements of performance in terms of distinct dimensions

The term "measurement of performance" is by itself vague, as there are several different dimensions of the performance on which one can focus. Gabrielsson (1999) sub-divides performance measurements into several categories as seen in Figure 2.3, with the most commonly measured dimensions of performance being *timing and dynamics*, and *intonation/vibrato*. In this section we will focus on the four dimensions studied in the dissertation rather than other types of performance measurements that focus on structure, conductors, and the performers' intention.

Timing and Dynamics

By the term *timing*, we refer to how the duration of single notes or other entities deviate from a norm, such as a "mechanical" performance with absolutely constant tempo and strict adherence to the ratios between note values in the score. *Dynamics* refer to the intensity of the produced sound, although between publications the measured property can vary from audio signal intensity to the intensity of the excitation mechanism, or even the listener's perception of musical dynamics (*piano*, *mezzoforte*,*forte*, et cetera).

Studying of timing and dynamics simultaneously has been the subject of empirical research for quite a long time, beginning around 1960 with the research efforts of Swedish musicologist Ingmar Bengtsson and his team (Bengtsson and Gabrielsson, 1977). Their research efforts were based around the hypothesis that music performance includes *systematic variations* in timing and dynamics, which they applied to a variety of musical styles (Viennese waltzes, Swedish folk music, piano performances of classical era pieces, et cetera). Almost all studies showed that "mechanical" (or perfectly robotic and "in time" performances) were practically non-existent even when the musicians were instructed to perform this way, and that there were certain commonalities in the variations of timing and dynamics across different performers. However, consistently large individual differences were also shown between performers for the same piece in various studies (Gabrielsson, 1987; Palmer, 1989).

Additional studies by Repp (1992, 1995, 1996) showed consistent timing and dynamics variations across repetitions from the same performer, and noticeable differences in timing variations between different performers. Individual differences in terms of dynamics variations were found to be less salient and only weakly related to individual differences in terms of timing variations. Another type of systematic variation called "melody lead", referring to notes in the melody being played slightly earlier than other concurrent notes, was investigated by Goebl (2001) and found to be a result of emphasis in dynamics combined with the mechanics of piano performance rather than intentional timing deviations.

Investigations on timing and dynamics continued with Todd's (1992) approach concerning a coupling between the two. Todd proposed a model to couple dynamics and timing, in which sound intensity is proportional to the square of musical velocity (number of events per unit time). Although the initial approach of considering a simple positive correlation between the two is now viewed as inadequate (Clarke and Windsor, 2000), further studies have found it difficult to pin down their relationship (Repp et al., 2002), while others debate the weight that is traditionally attributed to them.

However, they still remain ubiquitous and are probably the most studied performance dimensions, and have been used as a visualization aid in intuitively perceiving expressiveness in music performance (Dixon et al., 2002; Langner and Goebl, 2003) as well as the base dimensions for the extraction of expression patterns in piano performance (Widmer et al., 2003).

Intonation

Studies on intonation largely fall in two categories, each focusing on a different aspect: *vibrato*, and the choice of *tuning temperament*. Vibrato refers to a musical effect most commonly employed in vocal performance and bowed string performance, the modulation of pitch in a pulsating, sinusoidal pattern. Tuning temperament refers to fine adjustments to the intervals between notes in a musical scale, the most typically employed tuning temperaments being the *twelve-tone equal temperament* (i.e., identical intervals between subsequent notes) and *just intonation* (intervals defined by the harmonic series) among several other temperament systems.

Regarding vibrato, a large amount of the literature originates from research carried out at the University of Iowa under the supervision of C.E. Seashore in the 1930s. Most studies were based on empirical measurements of vibrato rate (frequency) and intensity (amplitude); average values were found to be 6.5 Hz and 0.6 of a whole-tone step, respectively. These values should be taken as an indication, as certain variation in both was found to be critical to the artistic quality of the vibrato (Seashore, 1932). Contemporary studies by Papich and Rainbow (1974) reported lower values for the vibrato rate in cello and double bass (around 4-5 Hz). Beyond quantitative descriptions of vibrato, Fletcher and Sanders (1967) demonstrated how vibrato can also cause continuous changes in the audio spectrum and therefore timbre, while Meyer (1993) reported similar results for dynamics.

The results of studies on choice of temperament and the performance of different intervals are less easy to summarize. Although there have been general observations such as an average variation of 40-60 pitch cents in the tuning of a given interval (Shackford, 1961) as well as common modifications on select intervals (Fyk, 1993, 1995), studies point out that intonation is a dynamic process that varies according to various factors including the expressive intentions of the performing musician.

Timbre

Timbre (referred to in Figure 2.3 as *perceptual effects*) remains an elusive concept whose perceptual and quantitative attributes have been the subject of study from many academic viewpoints (Krumhansl, 1989; Hajda et al., 1997). In contrast to other performance dimensions, timbre cannot be easily studied individually as it is inextricably connected to both dynamics and pitch (Krumhansl and Iverson, 1992). One of the most popular approaches to studying timbre perception is based on rating the dissimilarity of pairs of musical instrument sounds that primarily differ in their timbre (Peeters et al., 2011). This technique is most commonly applied in audio analysis and music information retrieval where the focus is on low-level spectral qualities of individual sounds, and thus it is not easily applicable to entire performances of musical compositions, the analysis and comparison of which is not so straightforward.

An alternative to relying solely on the produced sound as an insight to timbre, is using information on the way sound is produced. For the case of the violin, Schelleng (1973) demonstrated how bow-bridge distance is a parameter that is instrumental in the achievement of different timbral qualities in bowed string performance, with later studies expanding on bowing control parameters such as bow force and velocity in order to study the timbre of the produced sound to great effect (Maestre, 2009; Perez Carrillo, 2009; Guettler et al., 2003; Rasamimanana et al., 2006).

2.1.1.2. Measurements of the performer's movement

The importance of movement in music performance is undeniable, both as the main mechanism that produces sound, as well as an important channel for communication and expressivity. In the last two decades, advances in motion capture as well as video analysis technologies have allowed researchers to measure the movement of performing musicians in a highly accurate yet unobtrusive way, providing valuable insights into the influence of motor control and musical movement on performance. In her recent literature review on Music Performance, Palmer (2013) outlines two main functions of musical movement: *movement as sensory information*, and *movement as expressive gesture*.

Regarding movement as sensory information, the main hypothesis is that certain qualities of musical movement - such as finger movement in piano performance (Goebl and Palmer, 2008) or wind instrument performance (Palmer et al., 2009) - are essential sources of feedback that guide and enhance temporal coordination. Findings on piano performance showed that exaggerated movements that provide salient tactile feedback, such as striking a key with increased finger acceleration, help musicians improve the temporal accuracy of the subsequent note - thus providing an important aid in timing. This phenomenon was also reported more often for faster tempi, where temporal accuracy is arguably more difficult to achieve. Besides tactile feedback, Repp (1999) argues that visual feedback in solo piano performance is perhaps the least impactful on motor coordination – excluding the more challenging cases where the performer is either learning to play, executing large displacements of the hands, or sight reading. Motor feedback (in the form of proprioception) has been proven much harder to study and understand, primarily because of the strong coupling between auditory and motion feedback (Loehr and Palmer, 2009; Howell, 2004).

Regarding movement as expressive gesture, Davidson (2001) demonstrated various functions of body movement in vocal performance, including the communication of expressive intentions, exposing elements of the performer's character to the audience, and providing cues to co-performers. Individual studies by Wanderley (2002) and together with other collaborators (2005) showed how specific patterns of ancillary gestures, gestures that do not contribute to the generation of sound, were likely to repeatedly occur in the same part of the score for repeated performances, which demonstrates that they form an important part of the performance.

2.1.2. Ensemble music performance

As it can be evidenced by the way most literature review articles on Music Performance are structured so far, Ensemble Performance has not been traditionally considered as a standalone sub-topic within the Music Performance literature; this is especially reflected not just in the organization of sub-topics, but also in the volume of research carried out until the beginnings of the 21st century. In Gabrielsson's original literature review article (Gabrielsson, 1999), 9 out of 540 cited papers deal with ensemble performance, with 4 more papers having musical ensembles as their subject but coming from fields that are not typically related to music performance (such as Sociology). In Gabrielsson's 2003 updated version of the 1999 article (Gabrielsson, 2003), 4 out of 295 cited papers deal with ensemble performance, while in Palmer's review (1997) 2 out of 149 cited papers deal with Ensemble Performance, both of which overlap with the ensemble performance articles cited in Gabrielsson's reviews. Of course, these numbers do not represent the entirety of Ensemble Performance literature, but merely indicate the lack of a community dedicated to ensemble performance until the early 2000s.

In the last 10 to 15 years, a number of new research articles on Ensemble Performance have been published, either building on the case of individual performance or by studying ensembles from a new standpoint altogether. In this subsection we will present the ensemble performance literature most relevant to this dissertation, covering (i) synchronization, (ii) joint expression, (iii) research that goes beyond timing in ensemble performance, and (iv) social factors.

2.1.2.1. Synchronization among performers

Perhaps the reason why ensemble performance studies are under-represented in the literature is that it is easier to study a phenomenon in isolation for the case of one musician before moving on to its collective form. An example of this can be seen in studies on *Sensorimotor Synchronization (SMS)*, the coordination of action (e.g. sound-producing movements) with an external rhythm. SMS lies at the heart of ensemble performance as a synchronization mechanism, as musicians depend on the audible and visible actions produced by other members of the ensemble in order to coordinate their own; however, a review on SMS (also called the *tapping literature*) by Repp (2005) that covers some 280 papers predominantly presents work that is based on experiments with single participants, finger tapping along with the beat of a metronome or other pre-recorded auditory stimuli.

Among the many findings gained through research on SMS, two important synchronization phenomena are particularly interesting for our research: *phase correction* and *period correction*. These two mechanisms account for short-time and long-term adjustments to an external rhythm, respectively. Phase correction is typically viewed as an automatic process, while period correction is considered as being under cognitive control. Repp states that "although there may be alternative ways of conceptualizing this dichotomy, the dichotomy itself seems to be a robust finding". Mathematical models that approximate the two processes have been applied to real performance data several times in the past (Repp et al., 2012).

In a recently updated version of review on SMS, Repp and Su (2013) review new research directions taken since 2006, among which is the "collective form" of SMS (interpersonal sensorimotor synchronization). A fundamental difference between traditional SMS experimental setups and setups oriented towards ensembles is the fact that the external source of tempo is not a machine but a human, making synchronization bi- or multi-directional instead of unidirectional. Repp reports that “mutual entrainment among participants may occur and may not only facilitate entrainment to the external rhythm, but also make the task more enjoyable”. The results of several findings reinforce that last point, showing how tapping along to a human partner can facilitate synchronization as well as strengthen group cohesion (Kirschner and Tomasello, 2010; Wiltermuth and Heath, 2009).

Synchronization in an ensemble typically involves more complex temporal interactions than the ones observed in a rhythmic tapping task. Experiments under more natural performance conditions have been carried out by Goebl and Palmer (2009), who demonstrated the importance of both auditory and visual feedback in the synchronization in piano duet performance as well as how the absence of one type of feedback affects the musicians’ reliance on the other (additional research on the importance of visual and audio feedback in piano duets as well as piano-violin duets was carried out by Bishop and Goebl, 2015). Besides controlling the existence and type of feedback, the authors also assigned musical roles (leader/follower) to the musicians. The findings showed measurable effects of the musical role on the performance mainly for the case of the follower, who seemed to adjust their timing more than the leader and demonstrated overall less temporal accuracy. A cross-correlation analysis of the time intervals between successive notes showed that both musicians tended to follow each other in the case where full auditory feedback was provided, which together with comments by the participants suggested that cooperation is preferred to strict imposed musical roles.

Empirical work done by Moore and Chen (2010) on string quartet synchronization seems to provide additional evidence that musicians do not prefer strict leader/follower roles during the performance. Focusing on two members of the quartet as they executed a rapid passage of same-duration notes in synchrony, they found strong evidence of coordinative interaction (in their words “interactive coupling”) between them by analyzing their bowing movements. The results showed that each player followed a different micro-timing pattern that reflected different groupings of notes, and did not encounter evidence for one individual performer being more responsible for the achievement of synchrony than the other. In their words: “In the complex performance studied here every response is conceivably a stimulus, every stimulus conceivably a response; but neither can be identified”.

A lack of strictly assigned roles in an ensemble does not necessarily imply that all of its members equally influence its synchronization. In a comprehensive data-driven study carried out by Wing et al. (2014), two different string quartets were asked to perform a short excerpt while introducing intentional, unplanned expressive deviations in terms of timing in their performance. The authors utilized a linear phase correction model to quantify the amount of effort employed by one performer to adjust to the tempo of others, showing different results for each quartet. In one quartet, the first violinist exhibited less adjustment to the others than vice versa, while in the second quartet, the levels of correction by the first violinist matched those exhibited by the others; a result that highlights how quartets can employ different strategies ranging from an 'autocracy' to a 'democracy' when it comes to tempo coordination.

2.1.2.2. Joint musical expression

As we mentioned in Chapter 1, musicians playing together must not only achieve synchronization but also align their expressive intentions to jointly shape the performance. Keller (2008) formulates the previous statements as two requirements for joint musical action: (i) sharing common goal representations of the ideal sound, and (ii) possessing a suite of *ensemble skills* that enable these goals to be realized. Through several articles (Keller, 2008; Keller and Appel, 2010; Pecenka and Keller, 2009), Keller et al. describe and study three of these *ensemble skills* as the following cognitive mechanisms: *auditory imagery*, where the musician has his/her own anticipation of their own sound as well as the overall sound of the ensemble, *prioritized integrative attention*, where the musician divides his/her attention between their own actions and the actions of others, and *adaptive timing* which mainly deals with timing error correction mechanisms described in the SMS literature presented above.

Regarding the alignment of interpersonal expression and shared goals of the ensemble, the literature has markedly less contributions to offer, although the separation of "expressive" and "non-expressive" aspects of music performance is not a trivial or necessarily objective task. Studies on both solo and ensemble performance have shown that the structure of the performed composition (phrase boundaries, metric locations, rhythmic grouping) significantly influences expressive timing deviations (Repp, 1992; Rasch, 1988). Schögler (2000) found increased synchrony in improvising duets a few seconds before what he refers to as "points of change" in the improvisation, observing intense communicative interaction in order to carry out a subsequent change in the "feel" of the music - such as a transition from a tense, agitated section to a more relaxed section with a *legato* feel.

Goodman (2002) showed that gradual tempo changes (such as *accelerando* or *ritardando*) are more moderate during ensemble performance compared to solo performances of the same part. In an article on the temporal coordination of string quartets, Marchini et al. (2012) studied timing on two levels: *macro-tempo* and *mirco-tempo*, referring to tempo changes over a long (phrase level) or short (note level) timing window. The authors report similar results to the ones by Goodman (2002) in terms of macro-tempo, while the opposite was seen on a micro-tempo scale: ensemble performers would exaggerate the differences between rhythmic durations presumably in order to facilitate the communication of expressive intention. Similar results were reported on upper body and head movement, where sound-producing movements may be exaggerated and ancillary movements such as head gestures simplified (Glowinski et al., 2010; Goebl and Palmer, 2009).

2.1.2.3. Beyond timing: Intonation, Dynamics, Timbre

Within the context of ensemble performance, the number of published studies on other performance dimensions such as intonation, dynamics or timbre is relatively small compared to studies on temporal coordination and synchronization. One of the few related studies regarding intonation and timbre deals with singing in barbershop quartets (Kalin, 2005), where the results strongly suggested that the singers strive to separate their formants from one another in order to facilitate correct intonation; which also suggests that modulations in one dimension of the performance (in this case timbre) might be employed to achieve optimal coordination in another (intonation). A methodology for analyzing polyphonic vocal performances has also been outlined by Devaney and Ellis (2008), although, to our knowledge, it has not been yet applied in a published study.

Nickerson (1949) studied intonation in string quartets, reporting that Pythagorean temperaments are usually adopted in ensemble performance (which was also seen for solo performance as well), while Mason (1960) showed how ensemble musicians tend to choose the same temperament system. While differences between solo and ensemble performance have been observed for dynamics (Goodman, 2002), it is still unclear what their function is within the context of musical interaction; Goodman states that “musicians do not necessarily forfeit their own ideas when performing in an ensemble: the individual’s role is one of negotiation, or give and take”.

2.1.2.4. Social factors

Finally, regarding the social factors of performing in a musical ensemble, a study by Davidson and Good (2002) focused on both social interactions between the members of a student quartet as well as their temporal coordination, showing that collaborative activities between them depended on both in almost equal measure. They were able to observe a number of socio-cultural and socio-emotional issues among the quartet, such as concerns about performance anxiety, the overall quality of the performance, as well as imbalances in the inter-personal relationships of the quartet’s members. They also identified some cues important to coordination such as direct eye contact, shift of position to facilitate eye contact, as well as ”nods” from one performer to the other in order to show approval of a well-executed passage.

Williamon and Davidson (2002) recorded a series of videos from the rehearsals and final performance of a piano duet, and then jointly analyzed the material together with the performers. The results showed how rehearsals were used to consolidate the timing, phrasing and sense of musical style, while a combination of non-verbal gestures and eye-contact were developed in order to coordinate and communicate musical ideas. An extended study on the differences of body language between solo and ensemble performance in string quartets was carried out by Glowinski et al. (2013). Focusing on upper body and head movement, their relation to expressivity, and the purpose they serve in coordinated action, they highlight the challenges of carrying out empirical research on realistically complex scenarios.

2.2. Expressive performance

Research on “expressive performance” studies the deliberate deviations from the musical score that a musician performs in terms of different musical dimensions (intonation, dynamics, timing, timbre), as a way to express their own interpretation of the composition. The focus here is on the word *deliberate*: after all, even if a piece is performed by the same musician under the same circumstances twice, the resulting performances are bound to differ to a certain degree since it is nigh impossible for a musician to replicate every single action with perfect accuracy. In the previous section on Music Performance literature, we have seen that this very phenomenon (also referred to as *expressive deviations*, *systematic variations*, et cetera) has been studied since the beginnings of the 20th century.

What separates this field from the aforementioned music performance studies is the goal behind the analysis, as well as the methodology that is commonly applied for the description of the performance and the underlying score. The goal behind expressive performance studies is usually the creation of predictive models that, given an input score, are capable of introducing expressive deviations to a performance in a way that mimics a human musician (Widmer and Goebl, 2004; de Mantaras and Arcos, 2002). This is not necessarily done for the purpose of reproducing artificial performances, but often to understand human expressivity by observing the parameters of the trained model as well as the results it produces (Widmer et al., 2003). Regarding the commonly applied methodologies to expressive performance analysis, we have already repeatedly commented that a ubiquitous problem in music performance research is the volume of data generated during a single performance; expressive performance studies usually address this problem by utilizing computational analysis methods, as well as statistics to transform low-level performance data (audio samples, position and orientation data from motion capture markers) into mid-level performance *features* or *descriptors* to provide reduced information that is much closer to a musical context.

2.2.1. Expressive performance description

What is of special interest to this dissertation is the *performance description* aspect of expressive performance, since it is also our intention to apply such a methodology to our research. For the remainder of this section we will therefore present the most important examples of expressive performance studies to the extent that they deal with performance description.

2.2.1.1. Performance description from data acquired directly from the instrument

Detailed measurements of instrument data have been carried out for nearly a century, starting with the optical *Piano camera* system developed by Seashore's team at the University of Iowa (Henderson, 1936), which was capable of measuring note offset and onset times, key velocities, as well as pedal movement; a similar yet less obtrusive system called the *Photocell-equipped Bechstein piano* was designed and used by Shaffer (1980). With the advent of digital instruments and the MIDI protocol specification (Miles-Huber, 1991), researchers were able to directly capture digital representations of the performance simply by hooking up the instrument to a computer. It is important to note here that such instruments were not restricted to keyboards, but also included percussive devices (such as drum pads, xylophones, etc). Finally, the *Yamaha Disklavier* and *Bösendorfer SE* systems that appeared in the mid-80s offered the possibility of directly controlling an acoustic grand piano (as well as measuring detailed performance parameters) using a computer; besides a data acquisition tool, these instruments were also used as a way to play back synthesized performances using a real piano for the purposes of listening tests and evaluation.

The special importance of these systems as tools for analyses of expressive performance is the fact that the measured data are acquired on a note-by-note basis and can thus be contrasted on that same basis with the underlying musical score, using a digital symbolic representation such as MIDI files, or more recently MusicXML files (Good and Others, 2001). By comparing the acquired timing and dynamics values with the values specified in the digital score, the extraction of expressive deviations becomes a much simpler task.

Performance description extracted directly from the instrument becomes a much harder task for instruments that are continuously excited, such as bowed string instruments which are the focus of this thesis. Continuous excitation implies continuous control over the expressive parameters of the performance, which is a different paradigm from the case of keyboards and percussion. Anders Askenfelt (1986, 1989) was one of the first to acquire bowing motion descriptors (bow force, bow velocity, bow-bridge distance) during performance, using diverse custom electronic devices. More recent approaches (Rasamimanana et al., 2006; Demoucron, 2008; Schoonderwaldt and Demoucron, 2009; Maestre, 2009) improved the accuracy, intrusiveness and robustness of the sensing technique; but perhaps more importantly introduced computational techniques to encode, statistically summarize, model and predict the extracted performance descriptors, as well as relate them to the underlying score (in terms of dynamics, articulations, et cetera).

Bowing descriptors capture characteristics of the performed sound (such as timbre) which are either very difficult or impossible to indirectly extract from an audio recording (an equivalent example could be the use of pedal for piano performance). Within the expressive performance paradigm, they have been used to synthesize artificial performances using physical modeling (Maestre, 2009) and sample-based techniques (Maestre et al., 2010), as well as to perform timbral transformations on violin sound (Perez Carrillo et al., 2007; Perez Carrillo, 2009).

2.2.1.2. Performance description from audio recordings of the performance

Data acquired from the instrument are extremely valuable when the focus is on how the instrument itself was played, but there are still several important aspects of the performance that can be extracted from an audio recording of the performance and then contrasted to the underlying score in order to gain higher-level information on the performance. There is an entire field of research devoted to the musically meaningful analysis of audio signals, referred to by several different terms (Automatic Music Analysis, Music Processing, Music Informatics) but most commonly referred to as *Music Information Retrieval* or *Music Information Research* (MIR) (Serra et al., 2013).

Techniques from Music Information research can be typically used in expressive performance in two ways: as *pre-processing* tools to be applied before the performance description step, or as *extractors of musical features* that can be used to describe the performance. We will briefly go over the sub-topics involved within the MIR literature for each of these cases; for a comprehensive account of the presented topics see a recent book by Müller (2015).

Pre-processing
Source Separation deals with the decomposition of a sound mixture into its individual components (for a state-of-the-art review article see Vincent et al., 2014). For ambient recordings of musical ensembles where the individual audio of each instrument is not available, these techniques can be useful in obtaining the audio of each instrument so they can be studied independently from each other.

Music Structure analysis techniques aim at dividing a given music representation into temporal segments that correspond to musical parts, and to group these segments into musically meaningful categories (for an overview of the field as well as one of the most recent approaches see Nieto, 2015). We have already seen how the underlying structure of a composition influences expressive deviations; structure analysis is therefore a significant tool in carrying out computational analyses of expression in music performance.

Feature extraction
Spectral analysis techniques use the Fourier transform to obtain a frequency-domain representation of the audio signal, as a basis for extracting descriptor features for several different dimensions of the performance – the most typical being timbre descriptors, which are extensively used in studies of music perception (McAdams et al., 1995; Peeters et al., 2000); for a detailed listing as well as the function of each descriptor see Peeters et al.,2011). Spectral representations of audio are also widely used to extract pitch and melody information (Salamon and Gómez, 2012), although pitch is also commonly extracted from the time-domain representation of audio (De Cheveigné and Kawahara, 2002).

Music Synchronization refers to temporally linking different representations of musical content. The process of linking the audio representation of a musical piece to its underlying score is called *Score Alignment* and yields a temporal mapping which, for a given position in the audio recording, determines the corresponding position within a digital symbolic representation of the score (for an introductory article see (Dannenberg and Raphael, 2006)). This step is an often fundamental one in analyzing expressive performance, as it provides an immediate connection between score concepts (such as dynamics values and bar positions) and the recorded performance.

Tempo and Beat Tracking techniques deal with the automatic extraction of tempo and note onset information from an audio signal (an overview along with one of the most recent approaches can be found in (Zapata, 2013)). For recordings that are not based on an existing composition (such as improvisation) or where the score information is simply not available, such techniques can automatically extract valuable information on the timing dimension of the performance.

2.2.2. Ensemble expressive performance

To our knowledge, and beyond the related studies on ensemble performance that we already reviewed, there is scarcely any research applying expressive performance analysis and modeling techniques for the case of ensembles. Among the few examples, we can highlight the pioneering work carried out by Sundberg et al. (1989), who used an expressive performance modeling technique (called *analysis by synthesis*) to extract a set of rules for the automated performance of musical ensembles. A more recent study, inspired by Sundberg's approach, was carried out by Marchini (2014); the author proposes a novel methodology for building computational models of ensemble expressive performance, by introducing inter-voice contextual attributes (extracted from ensemble scores) and building separate models of each individual performer in the ensemble. Besides presenting a methodology to train predictive models of ensemble performance, the author also demonstrates how such techniques can be used to gain insights on ensemble expressivity.

2.3. Interdependence estimation methods

In Chapter 1, we stated that our goal is to define a methodology to measure interdependence between the members of a musical ensemble during a given performance, and outlined three methodological steps to achieving that goal: *performance capturing*, *performance description*, and *interdependence estimation*. So far we have reviewed the scientific background on music performance, placing focus on studies that deal with measurements of performance as well as methodologies used to extract descriptor features from the performance that are musically meaningful.

Numerical descriptors of performance are usually expressed along time, either as time series (such as pitch or audio intensity) or events (such as note onsets); in this dissertation specifically, we utilize extracted performance descriptors in the form of continuous time series. Several studies choose to represent these time series as the long-term statistical distributions of their local features, an approach known as the "bag-of-frames" (Aucouturier et al., 2007). In our approach we choose to not follow this paradigm, as we want to maintain the temporal relationship between these performance descriptors across performers. Therefore, estimating interdependence between the performers becomes a problem of estimating interdependence in multivariate time series; the last section of this chapter is dedicated to this very topic.

A great deal of research on multivariate time series interdependence has been carried out on neurophysiological signals, such as brain neuron activation patterns acquired using Electroencephalography (EEG); for an excellent review article we redirect the reader to Pereda et al. (2005), on which we have also based this subsection. Since the methods we introduce originate from different research fields, they vary in their complexity and required background knowledge. For that reason, we do not include a full mathematical formulation for each method; we refer the reader to the original publications where each method has been introduced for a complete description. We place our focus on measures that work with continuous signals, although the above referenced article presents additional methods that work with point processes.

2.3.1. Linear methods

Pearson product-moment correlation coefficient. This is the most commonly utilized metric for quantifying the linear dependence between time series x and y; its mathematical formulation describes the degree to which the means of x and y tend to vary in the same way, and is akin to a normalized version of the covariance between them. The output of correlation ranges from $\rho_{xy} = -1$, i.e. complete linear inverse correlation between time series x and y, to $\rho_{xy} = 1$, i.e. complete linear direct correlation between time series x and y, with a value $\rho_{xy} = 0$, suggesting an absence of linear dependence between the two time series.

Coherence. Linear interdependence between time series can be additionally assessed from their frequency representation; the coherence function calculates the linear correlation between two time series as a function of frequency, yielding a value between 0 (no association) and 1 (complete association) per frequency bin of the Fast Fourier Transform. This holds the advantage of being able to inspect the contribution of different frequency components of the time series towards interdependence, although coherence has been shown to be particularly sensitive towards both phase and amplitude relationships between the two time series.

Granger causality. In studying the relationship between variables, it is often useful to assess the *directionality* of that relationship; besides the overall degree of interdependence, it is also important to draw conclusions about the direction of influence, i.e. whether variable x is influencing variable y more than the opposite. A method that is capable of giving such an estimate is *Granger causality*, a statistical concept that was first applied to Econometrics (see Granger, 1969) and more recently to Neuroscience. It poses the hypothesis that if time series x causes time series y, then past values of x should significantly help in predicting future values of y as opposed to simply using past values of y to predict its own future; this is assessed through the use of a multivariate vector autoregressive modeling (or linear regression analysis, depending on the approach).

2.3.2. Nonlinear methods

Mutual Information. Mutual Information is a non-directional measure originating from the field of Information theory. It is not a *nonlinear* method per se, being based on the concept of *entropy* as proposed by Shannon in the 1950s and therefore dealing with reduction of information rather than the linearity of the relationship.

For a pair of time series x and y, mutual information measures the difference between two types of joint entropy; the joint entropy of the two variables as measured from the data, and the joint entropy of the two variables as if they were independent. If they are indeed independent, MI_{xy} is zero. Otherwise, MI_{xy} is a positive, non-bounded value that represents the amount of information that one gains about x by knowing the outcome of y.

Transfer Entropy is another measure originating from the field of Information theory. Rather than measuring the shared information between two variables, Transfer Entropy tries to compute the *flow* of information (or entropy) from one variable to the other, therefore providing directional information on the investigated interdependent variables. Introduced in (Schreiber, 2000), this measure has been used in the past to investigate causality between parts in musical compositions (Kulp and Schlingmann, 2009).

Nonlinear coupling coefficient. There exists a variety of nonlinear interdependence measures that quantify the signature of directional couplings between two time series $\mathbf{x}$ and $\mathbf{y}$; it is assumed that the processes behind the time series are characterized by separate deterministic dynamics which both exhibit an independent self-sustained motion. Assuming the existence of directionality, i.e. the dynamics of one process driving the dynamics of the other, directional couplings can in principle be detected by quantifying the probability with which close states of the driven dynamics are mapped to close states of the driving dynamics.

The state space of x (i.e. the set of all its possible states with each state corresponding to a unique point in the space of that set) is reconstructed using the method of *state space embedding*; then, a number of spatially nearest neighbors are selected for each point x_n in the state space (excluding temporal neighbors through the use of a given threshold). Finally, the squared mean Euclidean distance from the nearest k neighbors of x_n is calculated, along with the y-conditioned squared mean Euclidean distance (by replacing the nearest neighbors of x_n by the equal time partners of the closest neighbors of y_n). It has been shown that, when $x \rightarrow y$ coupling occurs, there is increased probability that close states of y are mapped to close states of x.

Several available measures based on the above paradigm exist; of these we highlight the measure L, which was recently shown to be of higher sensitivity and specificity for directional couplings than previous approaches. The output of $L_{x,y}$ is a bounded value between 0 and 1, with 0 signifying a complete lack of interdependence.For a more in-depth explanation of this particular method as well as a proper mathematical formulation, we direct the reader to Chicharro and Andrzejak (2009) where the method was originally introduced.

2.3.3. Method comparison

Table 2.1 shows a comparison of the various interdependence estimation methods presented in this subsection, summarizing them in terms of the following characteristics:

- **Type of interaction**, whether the method is fit for measuring interdependences of a linear or nonlinear nature
- **Directionality**, whether the method is not only capable of measuring the amount of interdependence but also inferring a direction of influence from one variable to the other
- **Number of variables**, whether the method assesses bivariate or multivariate relationships

Method	Type of interaction	Directionality	Number of variables
Pearson Correlation	Linear	No	Bivariate
Coherence	Linear	No	Bivariate
Granger Causality	Linear	Yes	Multivariate
Mutual Information	Nonlinear	No	Bivariate
Transfer Entropy	Nonlinear	Yes	Bivariate
Nonlinear coupling coefficient	Nonlinear	Yes	Bivariate

Table 2.1: Comparison of different interdependence estimation methods (rows), in terms of their characteristics (columns): the type of interdependence they can detect, whether they offer directional information on the interdependence, and the number of variables that they can simultaneously consider.

For most of the outlined methods there are variants that change their characteristics, such as adapting a method to account for nonlinear relationships, or to consider more than two variables simultaneously. The *Kendall* rank correlation and *Spearman* rank correlation coefficients account for nonlinear relationships by using rank statistics (Kendall, 1938; Spearman, 1904), while the Sample Multiple Correlation coefficient (Abdi, 2007) computes the correlation between one dependent variable and multiple independent variables using linear regression models. Similarly, there exist nonlinear adaptations of Granger Causality (Diks and Panchenko, 2006) as well as multivariate generalizations of Mutual Information (de Cruys, 2011). In this dissertation we choose not to delve into their characteristics and/or strengths and weaknesses, as we place the focus mostly on the basic concept behind each method and the types of problems they can solve.

Pereda et al. (2005) warn against bias towards the nonlinear methods, which can understandably exist as they may be perceived as more universally applicable and better equipped to deal with complex, real-life time series. The advantages and disadvantages of each method should always be informed by an empirical inspection of the studied time series in regards to their specific characteristics such as noisiness, stationarity, statistical distribution, et cetera. In conjunction to this, it is considered healthy practice to try out simpler, traditional linear approaches before moving on to more complicated ones.

Beyond the nature of the studied time series themselves, background knowledge of the phenomenon they represent is also essential to estimating interdependence. Background knowledge can allow us to perform pre-processing steps such as smoothing, detrending and removing certain frequency components from the original signal in order to eliminate influences that make interdependence estimation more difficult but do not otherwise affect the relationship between the studied time series.

Chapter 3

EXPERIMENTS

In this chapter we provide information on the experiments carried out throughout the duration of this dissertation. As we mentioned in Chapter 1, these consist of (i) an initial pilot study based on violin duets of both professional and amateur musicians, and (ii) a large-scale experiment on string quartets with a focus on carrying out research on musical interdependence as well as compiling an extended dataset that would be openly shared with the academic community at large. Consequently, we will first devote a short section to the pilot study and the obtained results, and then focus on the large-scale experiment, detailing (i) the experimental design, (ii) the selected musical materials and (iii) the technical setup. Finally, we describe the released dataset of experimental recordings.

3.1. Pilot study

An initial pilot study on the intonation of violin duets was carried out during the early steps of our research activities. Functioning as a stepping stone in developing our core methodology, this study served the purpose of making a first foray into studying interdependence in musical ensembles; in line with the exploratory nature of the study, we decided to target our research efforts on the relatively under-researched performance dimension of *intonation* in the context of an ensemble.

For the sake of conciseness and alignment with the rest of this chapter, we have chosen to mostly provide detail on the experimental design and conclusions. For a more in-depth description of the preliminary interdependence estimation methodology and/or a detailed description of the obtained results, we redirect the reader to (Papiotis et al., 2011) where a complete account of the study is provided.

3.1.1. Experimental design

We recorded two different violin duets performing two short pieces each. The first duet was composed of two professional violinists who already had previous experience in performing as part of a duet and were familiar with the pieces they were performing, while the second duet was composed of two amateur violinists who had no experience performing together and no explicit previous knowledge of the pieces to be performed.

Each piece was recorded in two discrete experimental setups:

- a *solo* setup, where each musician performed their part alone, and
- an *ensemble* setup, where the musicians performed their respective parts together as in a normal duet situation.

In order to reduce the complexity of the required task as well as motivate the musicians to focus on the performance without restrictions, the recordings were carried out without the use of a metronome. Prior to the recording of each piece, the performers tuned their instruments with a reference of A=440Hz. The pieces performed by the professional musicians were select excerpts from *Concerto for two violins* by J.S. Bach (BWV 1043) and *Duetti per due violini* by L. Berio. For the case of the amateur musicians, we opted for simpler scores with low difficulty; the pieces used were the traditional piece *Greensleeves* played in unison by the two violinists, and a simplified excerpt from *Duetti per due violini* by L. Berio. No specific instructions regarding the desired performance were given to the performers.

The sound of each violinist was captured using piezoelectric pickups fitted on the bridge of the violin, while a large diaphragm condenser microphone captured the overall sound of the duet. Besides audio signals, we also recorded violin bow motion data via a wired motion capture device (using the method proposed by Maestre et al., 2007), although they were not used in subsequent interdependence analyses.

In order to describe each performer's intonation behavior, we computed the deviation between the pitch produced by the performers and the pitch of the note they were performing; we then computed the Nonlinear Coupling coefficient (*L*-measure, outlined in Chapter 2.3.2 and proposed by Chicharro and Andrzejak, 2009) between each pair of pitch deviation time series as an estimation of the intonation interdependence between the performers. The Nonlinear Coupling coefficient was computed both for the ensemble and the solo conditions, and the obtained values from the two conditions were compared to each other in order to assess whether a measurable difference in terms of interdependence exists between the two conditions.

3.1.2. Brief description of results

Results from the computation of the nonlinear coupling coefficient were inconclusive, varying from duet to duet and piece to piece. The difference between ensemble and solo was clearest for the amateur duet, especially for the *Greensleeves* piece where both musicians performed the same melody in unison; for the multi-voice score of *Duetti per due Violini*, the difference between ensemble and solo was less pronounced. On the other hand, solo and ensemble interdependence for the case of the professional duet recordings was practically equivalent for both of the performed pieces. In order to interpret the inconclusiveness of the computational analysis and the difference in results between amateur and professional duets, we turned our attention towards a number of empirical observations made both during as well as following the recording sessions.

In all our recordings with the amateur musicians, three observations were made. First we noted that violinist 2, being less adept at sight reading, was more focused on performing his individual part of the piece correctly rather than listening to violinist 1. As a consequence, violinist 1 was mainly trying to adapt his intonation to violinist 2. Second, the piece performed in unison (Greensleeves) made the mismatch in intonation most audible, since the dissonance is much more apparent when the two violinists are performing the same melody, in the same tonal height; this recording also showed the clearest difference between ensemble and solo interdependence. Finally , the tempo of the performed piece seems to be another important factor; slow tempos exposed bad intonation, while fast tempos made it difficult for the musicians to keep their attention on their partner, presumably because of the cognitive load of the performance.

In the recordings with professional musicians our main observation was that, since the musicians were already familiar with each other's playing and the performed pieces, they demonstrated little difference in their intonation between solo and ensemble, and tended to shift their attention more towards the timing and dynamics dimensions of the performance. This became particularly evident when listening to the duet recordings, where it was nearly impossible to distinguish between ensemble and solo on grounds of intonation. Later informal listening sessions with other professional string instrument performers (Navarro, 2011; Donald, 2012) suggested that the intonation of the professional duet was subpar, hinting at a possible lack of intonation interdependence in the professional duet.

3.1.3. Conclusions

Intonation interdependence in violin duets proved challenging to observe, both from a computational point of view as well as from an empirical one. Save for the exception of the amateur duet performing in unison, it was not clear whether the challenge was due to our inability to measure intonation interdependence with computational means or due to the lack of a clear difference in intonation interdependence between the ensemble and solo recordings of the performances. The latter is clearly dependent on the experimental design of the pilot study, and thus we turned our attention to ways we can improve it.

In Chapters 1 and 2, we discussed how the existence of a shared goal is fundamental to achieving interdependence in ensemble performance. For the case of the amateur musicians in the pilot study, the potential lack of a shared goal could be attributed to a lack of skill, as we observed that one of the performers was directing most of his focus on performing the piece correctly without making individual mistakes. For the case of the professional musicians on the other hand, a more plausible explanation (given their skill level as professional performers) could be the lack of intent: the solo and ensemble performances were very similar to each other in terms of intonation, and other professional string instrument performers informally rated the overall intonation of the studied duet's performance rather poorly.

The above assume the lack of interdependent behavior in terms of intonation for the ensemble recordings. However, we must also consider the opposite case, where characteristics of interdependent behavior are present in the solo recordings as well; after all, results from the computational analysis of the amateur duet performances did show a difference between ensemble and solo interdependence for both of the recorded pieces. Contrary to the amateur duet, the members of the professional duet were already familiar with the pieces to be recorded. Thus, the common goals shared by the duet, if they did indeed exist, were already defined and beyond our control by the time of the recording. This implies that in the solo recordings, the musicians could be simply reproducing actions already learned during previous performances and/or rehearsals, or affected by mental imagery of the other member of the duet.

All of the above lead us to a series of conclusions, which where applied to the experimental design of our main experiment. First, the performers must be comfortable with the skill level required by the material to be performed. Second, it is necessary to conduct the experiments in a way that reinforces the need for a shared goal in the ensemble recordings. Finally, solo recordings should be carried out in such a way that limits any external influences normally present in ensemble performance to the bare minimum.

3.2. Main experiment

Upon receiving feedback on the design of the experimental conditions, we carried out our main experiment with increased complexity and depth in several aspects, to be detailed below. In contrast to the pilot study described above, in this experiment we captured the performance of a larger ensemble, a string quartet (two violins, viola and cello) of professional musicians.

The experiment was carried out during two research stays at McGill University in a collaboration between the following research laboratories: the *Music Technology Group (MTG)*[1] (Universitat Pompeu Fabra, Barcelona), the *Center for Computer Research in Music and Acoustics (CCRMA)*[2] (Stanford University, Stanford), the *Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)*[3] (McGill University, Montreal), the *Input Devices and Music Interaction Laboratory (IDMIL)*[4] (McGill University, Montreal), and the *International Laboratory for Brain, Music, and Sound Research (BRAMS)*[5] (Université de Montréal and McGill University, Montreal). MTG and CCRMA were responsible for the design and implementation of the experiments, while CIRMMT, IDMIL and BRAMS hosted and assisted in conducting the experiments.

The goals behind the preparation and realization of this experiment were the following:

- **To collect a ground truth dataset on string quartet interdependence.** In Chapter 1 we introduced the concept of *interdependence*, the coordinated variation of expressive aspects of the performance by the members of a musical ensemble in order to achieve a common aesthetic goal. We also explained how our main goal in this dissertation is to measure interdependence in terms of four distinct dimensions of the performance (intonation, dynamics, timbre and timing). The main goal of this experiment was therefore to obtain experimental material that represents the phenomenon of interdependence as clearly as possible, so we could design and test a methodology to measure it. To this end, we recorded several string quartet exercises, each with a clear shared interdependence goal.

- **To collect a ground truth dataset on string quartet expressive performance.** Interdependence itself is a means to the achievement of joint musi-

[1] http://mtg.upf.edu
[2] https://ccrma.stanford.edu
[3] http://www.cirmmt.org
[4] http://www.idmil.org
[5] http://www.brams.org

cal expression among the members of an ensemble. The field of *expressive performance* deals with the development of computational models that are capable of emulating musical expressivity, although virtually all of current work focuses on the case of individual musicians. As a first foray into building expression models for musical ensembles, a parallel goal to this experiment was the creation of a training dataset for computational models of ensemble expressive performance. To this end, we additionally recorded excerpts of string quartet pieces, where the shared goal of the ensemble was jointly shaped by the performers (rather than explicitly stated as in the case of the exercises mentioned above).

- **To share the collected data with the rest of the academic community for the purposes of further research.** Carrying out large-scale multimodal recordings with string quartets is a costly and technologically complex procedure. In the same way that we depend on the existence of quality experimental data in order to carry out our analyses, other researchers working on string quartet performance and related fields (ensemble performance, bowed string instrument performance, music information retrieval) can greatly benefit from having access to our experimental data, especially if it is infeasible for them to carry out experiments on a similar scale. Therefore, an overarching goal behind conducting our experiments was to capture data that would not only serve our research needs but also the needs of other researchers working on related topics. In the pilot study we captured individual audio from each performer, as well as bowing gestures captured with a wired motion capture system. For the main experiment we complemented the above with an ambient microphone as well as a binaural stereo recording device, a wireless optical motion capture system that additionally captured upper body movement, high quality video recordings, as well as questionnaire data.

In the remainder of this section we will provide in-depth information regarding the experimental design, the scores of each recorded excerpt, and the technical setup of the acquisition process. We will mainly focus on the material that was recorded in relation to the first and last goals presented above (collecting a dataset on string quartet interdependence, and sharing the collected data). For an in-depth account on the data that relates to the third goal (creating a dataset on string quartet expressive performance), we direct the reader to the doctoral dissertation of Marco Marchini (2014).

3.2.1. Experimental design

3.2.1.1. Overview

We aimed to capture a classical string quartet performing a series of musical tasks requiring the establishment of interdependence in order to achieve a shared goal. These tasks are divided in two groups:

- **exercises**, where the shared musical goal (e.g. a common *crescendo*[6] for all musicians) is explicitly stated along with the description of the exercise and the realization of that goal is feasible due to the simplicity of the exercise, and
- **excerpts of musical pieces**, where the shared musical goal is not stated but rather shaped throughout the rehearsals of the ensemble.

In order to use the recordings as ground truth data through which comparisons could be as unambiguous as possible, we centered our experiments around two conditions: *solo*, where each musician performs alone, and *ensemble*, where all musicians perform together as a group following a short rehearsal.

3.2.1.2. Materials

One of the pitfalls of the material selection stage is choosing scores which poorly or ambiguously demonstrate the studied phenomenon. For this reason, we consulted a string quartet expert (Navarro, 2011) who aided us in choosing the right scores, as well as in identifying which aspects of the performance are critical for achieving the shared musical goal in each score.

Exercises. We drew our material from the handbook "Exercises for String Quartet" by Mogens Heimann (1958). Quoting from the handbook:

> *The exercises for the string quartet form a toolkit for the art of chamber music playing and constitute a technique for sharpening collective perception in musical practice. Naturally, these exercises must be studied as a group project.*
>
> *Each exercise is a contribution to the expressive opportunities of collective music making. The author emphasizes the importance of focusing on one type of problem at a time, and each section deals with a particular topic.*

[6] A gradual increase in sound intensity over a passage according to western music notation.

Based on the handbook, we chose six sets of exercises, with each set concentrating on one aspect of the performance at a time. The sets are organized as follows:

1. **Intonation**, where the shared goal is to maximize the consonance of the overall sound by employing different tuning temperaments.

2. **Dynamic shading**, where the shared goal is to achieve matching dynamics values across different instruments, both in static dynamics (such as a simple direction to play *mezzoforte*[7]) as well as evolving dynamics (such as a crescendo).

3. **Rhythm**, where the shared goal is either the coordination of tempo changes or the execution of simple rhythmic patterns.

4. **Timbre**, where the shared goal is to generate matching timbres across different types of instruments (e.g. violin and cello).

5. **Unity of execution**, where the shared goal is to coordinate in terms of timing and dynamics so that the whole ensemble sounds "as a single instrument". This involves precise simultaneous execution of note attacks and releases[8], as well as smooth transitions in phrases that start in one instrument's score and continue in another.

6. **Phrasing**, where the shared goal is the coordinated execution of similar bowing techniques or articulations.

The exercises are grouped and described in a way which is more suited for a music practitioner rather than a researcher. For each one of these "performance aspects", we defined four distinct performance dimensions that we can associate them with (see Chapter 1.3). We carried out such an association for each exercise, based on comments by the string quartet expert we consulted (Navarro, 2011); it can be seen in Table 3.1.

---

[7]"Moderately loud" according to western music notation.

[8]In bowed string instruments, *attack* refers to the initial section of a played note, from the time the bowed string excitation starts until the note reaches its loudest point. *Release* refers to the final section of a played note, from the time the excitation stops until the note stops sounding.

	Intonation	Dynamics	Timing	Timbre
Intonation	X			
Dynamic shading		X		
Rhythm			X	
Timbre				X
Unity of execution		X	X	
Phrasing		X		X

Table 3.1: Association between categories of recorded materials (rows) and performance dimensions (columns).

The chosen exercises consist of short, simple musical tasks whose difficulty lies in achieving interdependence rather than correctly performing one's individual task successfully. Each exercise was accompanied by a set of instructions from the composer that were presented as the shared goal of the exercise (see Section 3.2.2); examples include *"the quartet should sound as one instrument"* or *"the crescendi and decrescendi must start at the same time and end at the same time."* Besides textual instructions, the Intonation and Timbre exercises also featured visual annotations that aided in comprehending the goal of the exercise, with the Intonation exercise featuring an upward/downward arrow denoting a departure from equal temperament to just intonation, and the Timbre exercise featuring a small visual aid denoting the distance of the bow from the instrument bridge; An example for each of these cases can be seen in Figure 3.1 ; detailed presentations of all exercises can be found in Section 3.2.2, while the scores for each exercise can be found in Appendix A.

Piece excerpts. Besides the musical exercises proposed above, we have also selected a number of pieces by well-known western classical composers. For each one of these pieces, we have isolated short excerpts with the help of IDMIL researcher Erika Donald (2012) that we can use as a more natural example of the phenomena demonstrated in the exercises.

Piece excerpts were chosen and categorized according to the six categories shown in Table 3.1 (Intonation, Dynamic shading, Rhythm, Timbre, Unity of Execution, Phrasing). The piece excerpt scores are more complex when compared to the exercises, and therefore inherently more challenging to perform. In order to ensure that the performers were comfortable with the piece excerpts, we contacted them a few weeks in advance to obtain a list of pieces from their rehearsed repertoire. The implications of this choice on the experimental conditions are discussed in the following subsection. Detailed presentations of all piece excerpts can be found in Section 3.2.2.

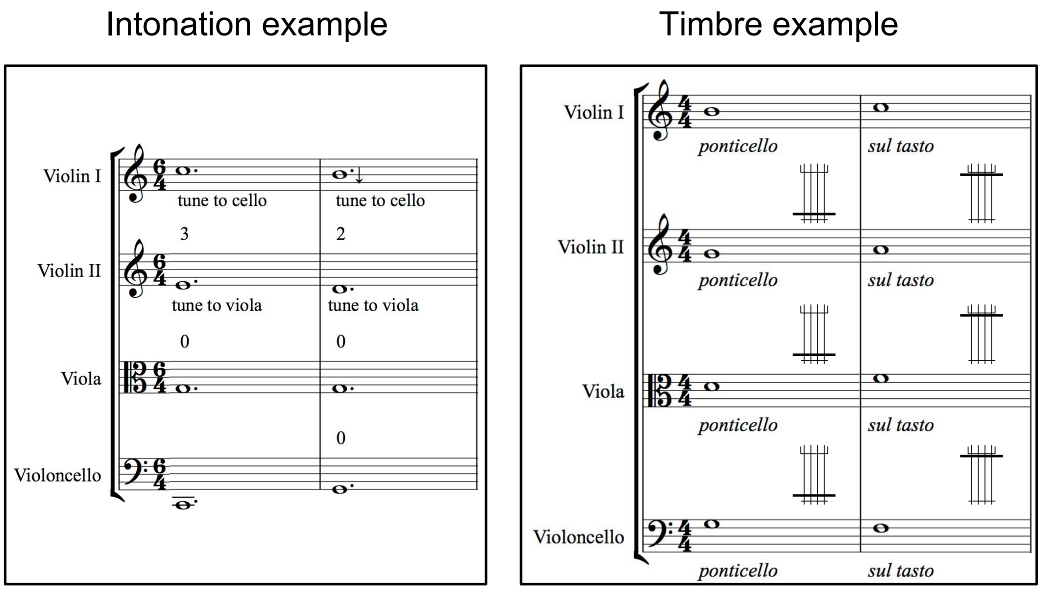


Figure 3.1: Examples of the visual annotations featured in the *Intonation* (left) and *Timbre* (right) exercises. In the Intonation exercise, upward and downward arrows signify that the note has to be played slightly sharper or flatter, respectively, while there are also textual indications as to which instrument should be used as an intonation reference. For the Timbre exercise, a small graphic is provided depicting the suggested bow position respective to the bridge and fingerboard of the instrument.

3.2.1.3. Conditions

The experiment was conducted with two main experimental conditions: *solo* and *ensemble*. Since there are methodological differences in terms of experimental design between exercises and piece excerpts, we will present the experimental conditions separately for each case.

Exercises. In the solo condition, each musician performed their part alone using a *stripped-down* version of the score; meaning they had no access to the scores of the other musicians nor the textual instructions or annotations therein. Essentially, the solo condition represents a case where both the shared goal and the communication channels between the musicians are absent. Each musician was provided with their solo version of the score prior to the actual recording, and was left to rehearse their part alone for approximately 15 minutes per exercise. Prior to each recording, we provided the musicians with four bars of a metronome click as an indication of the desired tempo, as well as a tuning reference of A=440Hz.

Following the solo condition, the musicians were given the complete quartet score along with the instructions of the composer in terms of the shared goal. They were then left to rehearse the exercise together for a short period (10-15 min) or until they felt they were achieving the shared goal satisfyingly. We did not interfere with the musicians during this rehearsal period, or attempted to control the rehearsal progress.

Finally, following rehearsal, the musicians were recorded in the ensemble condition; similar to the solo condition, they were provided with four bars of a metronome click immediately before recording as well as a tuning reference. For the case of the intonation exercise, we carried out two separate ensemble recordings: one without the annotations provided by the composer, and one with the annotations. For the case of the timbre exercise, the provided annotations were present in both the solo and ensemble scores since they referred to the individual playing technique rather than the common goal of the ensemble.

The exercises were recorded over the course of several days; all of the solo condition recordings were carried out during the first day, with the rehearsals and ensemble condition recordings spread out throughout the rest of the days. The consistent solo-ensemble ordering was a necessity of the experimental design, as the lack of familiarity with the scores of the rest of the ensemble is a core characteristic of the solo condition. This could potentially entail a risk of order effects in our subsequent interdependence analysis due to the performers being more familiar with the scores by the time of the ensemble recording through additional exposure to them; we consider the simplicity of the exercises as a significant reduction to that risk, and in fact the musicians did not seem to encounter any difficulty in performing their scores in the solo condition.

Piece excerpts. The musicians already had prior knowledge of the scores of each excerpt since they were part of the quartet's repertoire. Thus, in the solo condition each musician performed their part alone, while in the ensemble condition they all performed together following a short warm-up rehearsal.

This implies a fundamental difference between exercises and piece excerpts in terms of the two experimental conditions. The fact that the performers are already familiar with the full ensemble score and have already rehearsed with the ensemble in the past means that solo performance is potentially affected by *auditory imagery*, a mental anticipation of the sounds of the rest of the ensemble. This in turn implies that the piece excerpts are not comparable to the exercises in terms of their potential for low-level, ground truth analyses of interdependence; at the same time, they can be seen as a better fit for higher-level analyses of music ensemble expressivity.

In this dissertation, we carry out an analysis of interdependence based solely on the exercise recordings. For an analysis of ensemble expressivity using the piece excerpt recordings we direct the reader to the work recently presented by Marchini (2014).

3.2.1.4. Subjects

The subjects in the main experiment were all members of a professional string quartet ensemble based in Montreal, Canada. In order to guarantee the reliability of the captured data, we considered the following criteria when searching for candidates:

- **Adequate level of technical prowess**. The amateur recordings of the pilot experiment highlighted an important requirement in carrying out experiments on interdependence: the subjects need to be comfortable with the difficulty level of the performed score, lest they end up focusing too much on playing their own part correctly and ignoring the rest of the ensemble.
- **Experience in string quartet performance.** This is perhaps the most important quality that is needed of our subjects. Instrumentalists who do not have experience with quartets would not serve our purpose since they have no first-hand experience of the qualities that an ensemble must possess. According to the quartet expert we consulted (Navarro, 2011), even a combination of four world-class soloists can lead to mediocre results, if the musicians are not striving for interdependence.

- **Experience in performing with one another as a team.** To facilitate rehearsals, it is best to work with a quartet of musicians that are already used to performing with one another. This way, the social dynamics between them are already in a steady state.

The chosen subjects were members of a professional string quartet that had already been performing together for more than a year, including public performances. The mean age of the musicians was 30 years old ($\sigma = 2.9$), and all of them had been practicing their respective instruments for at least 20 years, including at least 10 years of professional experience as members of musical ensembles (chamber orchestras, symphonic orchestras, string quartets). They were all compensated for their participation in the experiment, and signed an agreement (jointly prepared by the participating academic institutions) regarding the use of their data.

3.2.1.5. Questionnaires

Following the recording of each exercise in the ensemble condition, all members of the ensemble were asked to fill out a questionnaire. The performers were asked to rate the following, using a Likert scale:

1. The difficulty of their own part of the exercise as a personal task
2. The difficulty of the exercise as an ensemble task
3. The degree of success with which the musician performed their personal task
4. The degree of success with which the ensemble performed the task
5. The clarity of the goal of the exercise
6. The importance of their part in achieving the shared goal
7. The degree to which they were focused on playing their part or on achieving ensemble cohesion

Additionally, each member of the ensemble was also asked whether they believed that a particular member of the ensemble had a "leader" role in achieving the ensemble task; if the answer was positive, they could specify which member of the ensemble represented that role. In our own computational analyses of interdependence, we do not explicitly study the relationships between the performers nor the existence of a leader within the ensemble. However, we decided to collect such data as it may prove important for future research using our dataset.

Clarifications on the questionnaire, where necessary, were given to the musicians *in situ.* The template used for the questionnaires as well as a preliminary analysis of the questionnaire data can be found in Appendix B.

3.2.2. Detailed description of selected scores

In this subsection, we provide details on each recorded exercise and piece excerpt. Even though in this dissertation we only utilize the exercise recordings for the purpose of investigating interdependence in string quartet ensembles, we provide in-depth descriptions for all recorded materials as documentation and a guide for other researchers that may take advantage of the recorded materials in the future.

All experiments are divided in six sets, following the structure of the exercise handbook as described in Subsection 3.2.1.2. For each set, we initially provide a brief introduction of the musical phenomenon that we wish to focus on. We then describe the recorded exercises and piece excerpts.

3.2.2.1. Intonation

In this set of experiments, we begin with an exercise on string quartet intonation. Quoting from Heimann (1958):

> *"Intonation" means much more than just playing in tune. A quartet which does not yet know this from experience will be able to broaden its horizons during intensive work on this section. As in all the studies, the necessary conditions for the success of the work are absolute concentration and intense listening.*

The goal of the exercise (**I1**) is not to train the musician to "play in tune", since this is already considered a necessary trait of the quartet members. Instead, the objective here is to employ different tuning systems in order to maximize the consonance of the produced chord. This is achieved by following the instructions annotated on the score: each musician is instructed on which instrument should be used as an intonation reference, and whether they need to slightly raise or lower the produced pitch in order to achieve good intonation.

The first piece excerpt on intonation (**IP1**) is taken from the fourth movement (*Alegretto*) of Beethoven's string quartet n.4 in C minor (opus 18)[9] and consists of bars 16 through 24, repeated twice. It consists of a slow-moving, largely isochronous four-part harmony, which the quartet was instructed to perform at a

[9]Available at http://imslp.org/wiki/String_Quartet_No.4,_Op.18_No.4_(Beethoven,_Ludwig_van)

significantly lower tempo (90 bpm) compared to the tempo at which the piece is usually performed (140 bpm).

The second piece excerpt (**IP2**) is taken from the fourth movement (*Vivace*) of Haydn's string quartet n.3 in Eb major (opus 71)[10] and consists of bars 117 through 124, repeated twice. Similarly to the previous piece excerpt (although performed at a faster tempo), it consists of an isochronous harmony which is ascending by semitone increments.

3.2.2.2. Dynamic shading

In this set of experiments, we aim to study how a string quartet handles dynamics both for instant as well as gradual changes in dynamics. In the first exercise (**D1**), the dynamics value assigned to each instrument is static (i.e. does not evolve with time). In the second exercise (**D2**), the quartet is tasked with performing coordinated *crescendi* and *diminuendi* starting from (and ending with) different dynamics values.

The first piece excerpt on dynamics (**DP1**) is taken from the first movement (*Allegro moderato*) of Borodin's string quartet n.2 in D major[11] and consists of bars 280 through the end of the movement. It consists of a long, drawn out *decrescendo* which lasts for 24 bars, interspersed with accents for the instrument that is playing the melody (initially the first violin and subsequently the viola).

The second piece excerpt on dynamics (**DP2**) is the same as the second piece excerpt on intonation (IP2). From a dynamics point of view, it features a long *crescendo* drawn out over the duration of the excerpt.

The third piece excerpt on dynamics (**DP3**) is taken from the first movement (*Allegro ma non tanto*) of Beethoven's string quartet n.4 in C minor (opus 18) and consists of bars 53 through 77. It features a particular pattern of synchronized accents played simultaneously by the entire quartet, requiring cohesion in terms of articulation. Additionally, the last seven bars feature abrupt alternations between quiet and loud dynamics (from *pianissimo* to *forte* and vice versa).

3.2.2.3. Rhythm

The exercises on rhythm are designed to train the ability of the musicians either to coordinate on tempo changes or to realize clear rhythmic patterns.

---

[10] Available at http://imslp.org/wiki/String_Quartets,_Op.71_(Haydn,_Joseph)

[11] Available at http://imslp.org/wiki/String_Quartet_No.2_(Borodin,_Aleksandr)

The first exercise (**R1**) is focused on carrying out coordinated tempo changes with rather simple rhythmic patterns, either adopting a slightly faster tempo (*poco piu mosso*) or a slightly slower tempo (*poco meno mosso*). The second exercise (**R2**) features a sequence of galloping rhythms based either in binary or in ternary subdivisions of the tempo. The objective is to realize those rhythms and with a clear distinction among them.

The first piece excerpt on rhythm (**RP1**) is taken from the first movement (*Allegro ma non tanto*) of Beethoven's string quartet n.4 in C minor (opus 18) and consists of bars 136 through 148, repeated twice. It features a prolonged section that requires tight synchronization at an eighth-note level between the second violin, viola, and cello, while the first violin is performing a solo.

The second piece excerpt on rhythm (**RP2**) is taken from the second movement (*Andante scherzoso*) of Beethoven's string quartet n.4 in C minor (opus 18) and consists of bars 1 through 32. It features a canon-like structure that requires temporal coordination in order to strike a balance between individual expression and temporal cohesion to make all voices "fit together".

The third piece excerpt on rhythm (**RP3**) is taken from the third movement (*Menuetto*) of Beethoven's string quartet n.4 in C minor (opus 18) and consists of bars 8 through 50, repeated twice. It features a 3/4 waltz-like passage where the musicians are required to play each note simultaneously for a prolonged period.

3.2.2.4. Timbre

In this experiment, we want to study how timbre similarity is perceived and achieved by the musicians of a quartet, by providing the musicians with specific instructions (bow positions) together with high level descriptions of the produced sound. We intend to see whether these instructions will yield the same results in tone/timbre similarity, and whether the musicians will alter their produced sound in light of the new specificity.

The first (and only) timbre exercise (**T1**) focuses on the effect of the bow-string contact point in conjunction with dynamics; each bar is concerned with the tone produced by a specific contact point and a specific dynamic level. Quoting from the exercise description given by Heimann (1958):

> *A descriptive name is given for each tonal quality or "sound color": for example, the sound color produced by playing mezzo-forte with the contact point at the instrument's bridge is called "ponticello", while if the contact point is slightly moved away from the bridge and*

the dynamic level is reduced to mezzo-piano, then the sound color is termed "oboe".

We did not record any piece excerpts with explicit focus on timbre, as we could not find a passage related to the T1 exercise within the quartet's repertoire.

3.2.2.5. Unity of Execution

Unity of Execution refers to making the entire quartet sound "as a single instrument", mainly in terms of high coordination regarding timing and dynamics. In the first unity of execution exercise (**U1**) the musicians are required to play a sequence of notes with precisely synchronized note attacks and releases. In the second (**U2**) exercise, two musicians at the time have to play together while the other two have rests, and the alternation between the duets has to be as smooth as possible.

The first piece excerpt on unity of execution (**UP1**) is taken from the fourth movement (*Allegretto*) of Beethoven's string quartet n.4 in C minor (opus 18) and consists of bars 24 through 40, repeated twice. It is similar to exercise U2 in the sense that seamless continuation is required between phrases performed by different instruments.

The second piece excerpt on unity of execution (**UP2**) is taken from the first movement (*Allegro ma non tanto*) of Beethoven's string quartet n.4 in C minor (opus 18) and consists of bars 1 through 25. It poses the challenge of maintaining "togetherness" while performing different simultaneous note durations and articulations, especially between the first violin and the rest of the quartet.

The third piece excerpt on unity of execution (**UP3**) is the same as the the third piece excerpt on dynamics (DP3).

3.2.2.6. Phrasing

The exercise on phrasing (**P1**) focuses on well-coordinated bowing control among musicians. It features a descending and ascending scale performed with *slurred* and *legato* articulations.

The first piece excerpt on phrasing (**PP1**) is taken from the third movement (*Andante*) of Borodin's string quartet n.2 in D major and consists of bars 111 through 133. It features groups of slurred notes (especially for the second violin) with different, overlapping phrase lengths between instruments.

The second piece excerpt on phrasing (**PP2**) is the same as the third piece excerpt on rhythm (RP3).

3.2.3. Technical setup

During the experiments, each performance was acquired using multiple data modalities simultaneously. The objective was to synchronously acquire, process, and conveniently store three different types of data: *audio* data captured using ambient microphones as well as contact microphones placed in the bridge of the instruments, *motion capture* data using two different high-end systems (one wired and one wireless), and *video* data using a high quality video camera.

A schematic illustration of the technical setup can be seen in Figure 3.2; in it, five major entities are shown, each in charge of a different function: (1) a synchronization signal generator, (2) a machine devoted to the acquisition of audio signals as well as the acquisition and processing of motion capture data using a wired system, (3) a machine dedicated to the acquisition and processing of motion capture using an infrared camera-based motion capture system, (4) a machine dedicated to the acquisition of high quality video, and (5) a machine serving as a data repository. A photo from the experiment recording sessions can be seen in Figure 3.3.

We will present the technology used for the acquisition of each data modality in the following subsections, as well as the synchronization protocol we implemented to temporally align the acquired data streams.

3.2.3.1. Audio capture

Both individual audio from each performer as well as ambient audio was captured in order to study the performance in terms of the produced sound.

Individual audio from each performer. An individual audio signal from each performer was acquired using a piezoelectric pickup fitted on the bridge of each instrument. We used *Fishman V-100* pickups for the violins and viola, and a *Fishman C-100* pickup for the cello. All recordings were carried out using a sampling rate of 44.1 Khz. The use of pickup microphones holds the advantage of capturing solely the sound of the instrument it is attached on and removing any sonic interference from the rest of the instruments or the room ambiance. Pickup gains were manually set with the aid of level meters in the recording equipment to avoid clipping of individual audios. To balance the sound level of the quartet as a

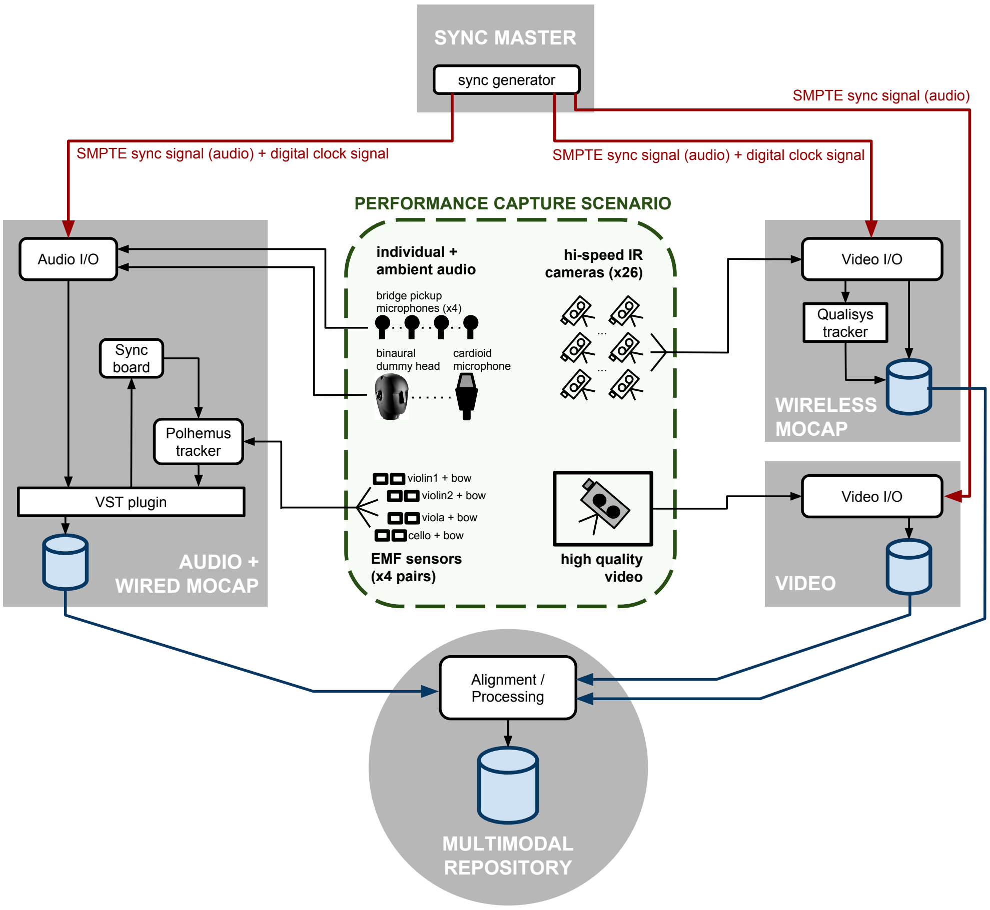


Figure 3.2: Schematic of the technical setup for the main experiment recordings. Red lines represent synchronization signals, black lines represent raw sensor data, and blue lines represent processed data to be stored.

Figure 3.3: A photo from the experiment recording sessions. The musicians' faces have been blurred out for privacy.

whole, we employed an iterative procedure, based on listening to the mix of the pickups on the headphones while adjusting the four gains.

Ambient audio. Ambient sound - i.e. the overall sound produced by the musicians - was recorded using two different devices, one monophonic and one stereophonic. For monophonic audio, we used a *Neumann KM184* cardioid condenser microphone sampled at 44.1 KHz. The placement of the microphone was approximately 2 meters away from the centroid of the semicircle defined by a quartet's typical seating arrangement, in the direction faced by the quartet. For stereophonic sound, we used a binaural dummy head (Neumann KU-100) placed directly below the condenser microphone, facing the quartet.

3.2.3.2. Motion capture

We used two different motion capture systems simultaneously to record each performance, a wired system based on electromagnetic field (EMF) sensing, as well as a wireless system based on optical tracking of infrared light-reflective markers. The systems capture two types of movement: bowing motion via wired sensors and markers attached to the body and bow of each instrument, and upper body motion via markers attached on the performers. Data acquired from the two systems overlap in the acquisition of instrument body and bow movement, and can be potentially used to compare the accuracy of one system versus the other.

Wired system based on electromagnetic field (EMF) sensing The first of the two motion capture systems used in our experiment is the *Polhemus Liberty*[12] system, a six-degrees-of-freedom (6DOF) tracking system based on electro-magnetic field (EMF) sensing. It consists of a transmitting source, and a set of receiving wired spheric sensors with a radius as low as 0.5 cm and a weight down to 6 grams. Each sensor provides 3DOF for translation and 3DOF for rotation at a 240 Hz sampling rate, with translation and rotation static accuracies of 0.75mm and 0.15 degrees RMS respectively within a range of 4m of distance to the source when using the *Long Ranger* source model.

Besides the standard wired sensors that provide position and orientation, a supplementary sensor resembling a stylus is used for manually annotating instantaneous positions. Using the tip of the stylus, the user can store the position of that point relative to the coordinate system defined by the wired sensor via the click of a button; this way, solid bodies for which deformation can be neglected are defined by attaching a sensor to the body and annotating its edges using the stylus.

The EMF-based tracking system was used to capture instrumental (i.e. sound-producing) gestures using the method described by Maestre et al. (2007). This method uses two sensors per tracked instrument: one wired sensor fixed on the back plate of the violin in order to track its position and orientation, and another sensor fixed on the bow near the frog; the exact placement of the sensors can be seen in Figure 3.4.

Using the stylus sensor, the positions of both ends for all four strings are marked in relation to the instrument sensor, while the four extremities of the bow hair ribbon (bow frog left/right, bow tip left/right) are marked in relation to the bow sensor; a visual explanation of the marked points can be seen in Figure 3.5.

Using these positions, we are able to compute the following instrumental gesture features:

- *Bow displacement* (also called *bow transversal position*), the distance between the bow frog and the bow-string contact point.
- *Bow-bridge distance*, the distance from the bow-string contact point to the instrument's bridge.
- *Bow inclination*, the rotation of the bow around its pitch axis.
- *Bow tilt*, the rotation of the bow around its roll axis.
- *Bow angle*, the rotation of the bow around its yaw axis.

[12] http://polhemus.com/motion-tracking/all-trackers/liberty

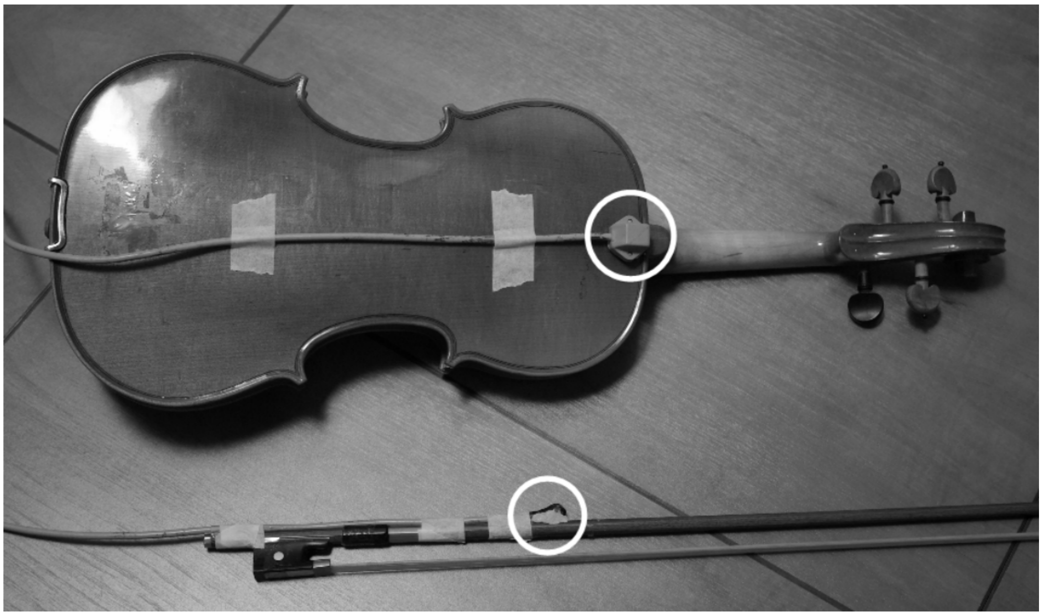


Figure 3.4: Detail of the wired motion capture sensors placed on the instrument and bow for the acquisition of instrumental gestures.

Figure 3.5: Detail of the marked string and bow ribbon positions relative to the wired motion capture sensors used for the acquisition of instrumental gestures.

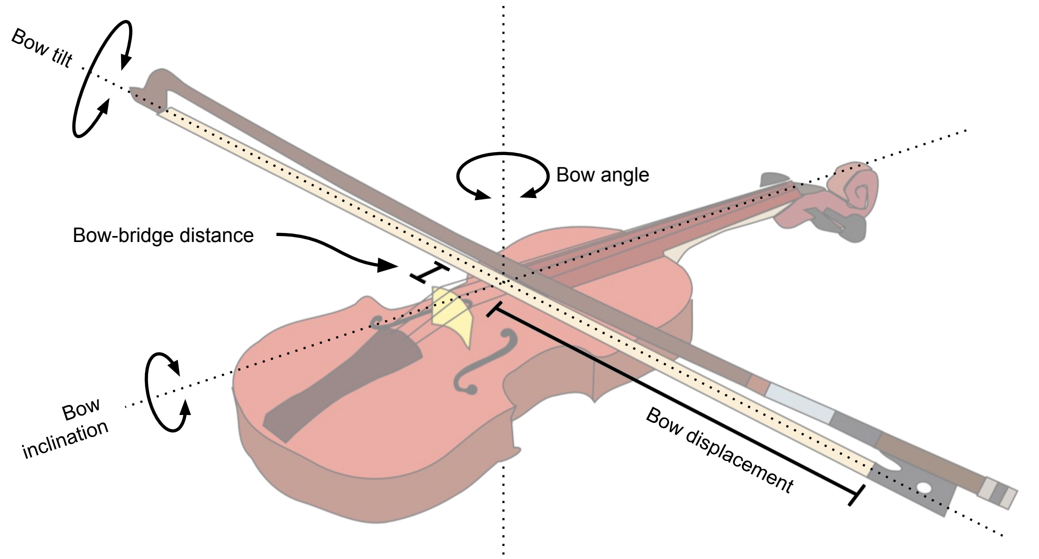


Figure 3.6: Schematic of the instrumental gestures that are computed from the relative position of the bow with respect to the instrument.

A visual schematic of the above features can be seen in Figure 3.6. From these features, we are able to additionally compute the following:

- *Bow velocity*, by computing the first-order derivative of the bow displacement feature.

- *Current string being played*, by measuring the angle of the bow inclination in respect to the plane defined by the instrument body.

- *Bow pseudoforce*, a rough estimate of the bow pressing force measured from the deflection of the bow hair ribbon as explained by Marchini et al. (2011).

The EMF motion capture data are recorded simultaneously with audio data in a *Virtual Sound Technology* (VST) software plug-in developed at the Music Technology Group (MTG) of Universitat Pompeu Fabra. Initially developed for the purposes of capturing motion data from a single violin (Maestre, 2009; Perez Carrillo, 2009) and extended for the case of a string quartet (Marchini, 2014), this plug-in not only records audio and motion capture data but additionally computes the instrumental gesture features described above in real time. A screenshot of the plug-in in action can be seen in Figure 3.7.

Wireless system based on optical tracking of infrared light-reflective markers. The second motion capture system used in our experiment is developed by *Qualisys*

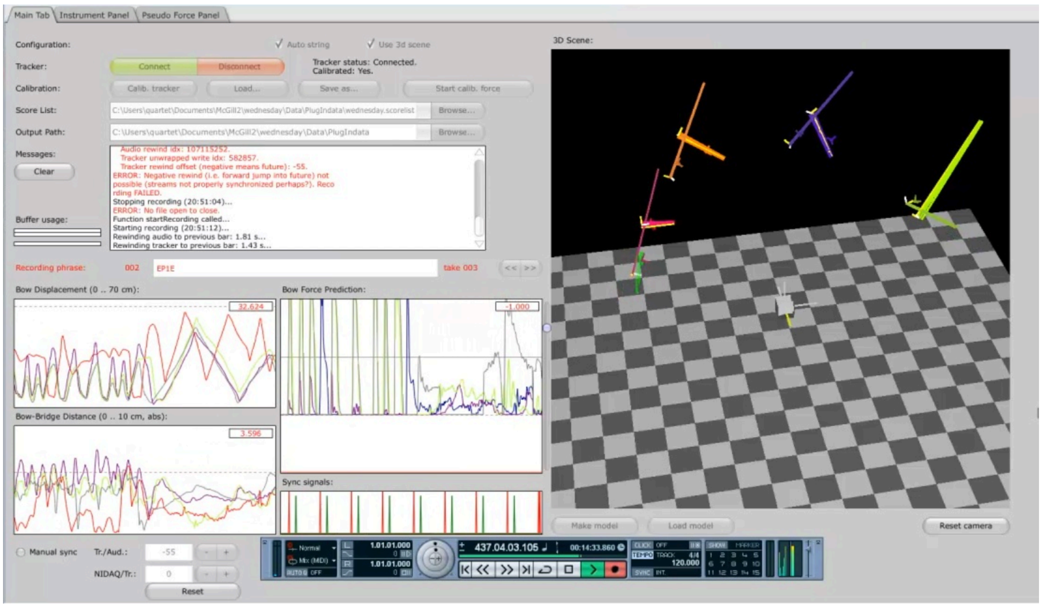


Figure 3.7: Screenshot of the VST plugin used to record audio and wired motion capture data. A 3D visualization of instrument strings and bows for the entire quartet can be seen on the right, while a various instrumental gesture features (bow displacement, bow-bridge distance, bow pseudoforce) can be seen on the bottom left.

Inc.[13]. It is composed of 32 *Oqus 400* infrared cameras, each one capable of 3 MegaPixels of resolution and of a maximum of 485 frames per second. For our experiment we used only 26 cameras since the space occupied by the quartet was small enough to cover using a subset of the available system. We spatially arranged the cameras in a way that minimized marker occlusion through trial and error. The focus and aperture of each camera lens was manually adjusted, together with the exposure time, to match the specific condition of each camera due to the non-uniform lighting of the room where the system was installed.

The infrared camera-based system was used to capture both instrumental gestures as well as upper body movement. Regarding the instrumental gestures we used the marker setup reported by Schoonderwaldt and Demoucron (2009); the reason behind using two different motion capture systems simultaneously to acquire instrumental gestures was to perform future analysis on their comparative accuracies. Regarding upper body movement, we used the standard marker placement protocol used by the IDMIL laboratory[14].

3.2.3.3. Video capture

High quality video was acquired with a *Canon VIXIA HF R200* camera; the audio track of the video was fed with a SMPTE linear timecode audio signal, used as a synchronization index to temporally align the different data streams. Further details on synchronization are provided in the subsequent Section 3.2.3.5.

3.2.3.4. Musical instruments

Since a variety of sensors (pickup microphones, wired motion capture sensors, reflective infrared markers) had to be fixed on the instruments and bows, it was not possible for the musicians to perform with their own instruments for fear of wear and/or damage to the finish or the instruments themselves. For that reason, we used four instruments kindly contributed by the *Schulich school of Music*[15] of McGill University.

[13] http://www.qualisys.com/

[14] Available at: http://www.idmil.org/mocap/Plug-in-Gait+Marker+Placement.pdf

[15] https://www.mcgill.ca/music/music

3.2.3.5. Synchronization

Several different devices were used to record the acquired data. A laptop computer with an external audio interface (*RME Fireface 800*) was used for both audio as well as wired motion capture data, a desktop computer was used for optical motion capture data, and a video camera for the high-quality video.

In order to minimize the effects of drift and/or jitter in the digital sampling clocks of the various recording devices, a "master" external clock signal source (*Rosendahl Nanosyncs HD*) was sent to both computers to control their respective sampling clocks. Additionally, a linear timecode signal (SMPTE[16]) was sent to the two computers as well as the video camera, in order to associate each captured frame to a common time reference.

The Qualysis system (used for optical motion tracking) allows for synchronous operation from an external clock and timecode source. However, the Polhemus Liberty system (used for wired motion tracking) only offers the possibility of detecting incoming Transistor-transistor logic (TTL) pulses (fed into its sync-in pin) and inserting a flag in the most recent data frame.

Using custom recording software developed in-house, the audio interface connected to the laptop computer would periodically send an audio pulse to a custom-built analog conversion board that transforms it to a TTL pulse, which was then fed into the Polhemus sync-in pin. Audio/Polhemus synchronization is carried out in real time, measuring the delay between the two signal acquisition devices by counting the received audio samples from the time the pulse was sent by the audio I/O interface, until the time the converted TTL pulse is detected as a data flag within the incoming Pohemus data frames. Fluctuations in the delay are compensated for by either discarding or repeating incoming data frames. Qualysis 6DOF data and Polhemus 6DOF data are then processed and stored with a common reference time.

An overview of the route of each synchronization signal can be seen in Figure 3.2.

3.3. Dataset

All of the recordings carried out as part of the main experiment outlined in the previous section are publically available as a dataset named QUARTET[17]. Prior to sharing the recordings, we carried out a temporal alignment between each recorded

[16]https://www.smpte.org/standards

[17]http://mtg.upf.edu/download/datasets/quartet-dataset

performance and its underlying score, obtaining the note onset and offset times for each note in the recorded performance; the following subsection explains the motivation and background behind this procedure, and provides detail on its implementation. Then, we present the repository platform through which the QUARTET dataset is made available, and provide an guide to the contents of the dataset.

3.3.1. Score-performance alignment

3.3.1.1. Background

The existence of an underlying musical score for each of the recordings provides a reference to which the performance can be compared. This comparison between performance and score is of especially high importance for performance analysis tasks and most significantly *performance description* (discussed in Chapter 2), where it is the main tool used to characterize the deviations of the performer in relation to the score. The first step towards carrying out this comparison is called *score-performance alignment* or simply *score alignment*; the aim is to associate events in the score such as note onsets with specific times in the audio recording of the performance.

This temporal association can be obtained in various ways. The most basic way simply involves listening to the recording - usually through an audio editor that provides additional visual feedback in the form of the recording's waveform or frequency domain representation - and manually noting down the exact times on which note onsets and offsets occur. There are several audio editors that are used to facilitate this process by offering manual annotation tools, the most typical being *Wavesurfer*[18] and *Sonic Visualizer*[19]. However, this is an arduous manual task that can require a time commitment that amounts to multiple times the duration of the aligned recording.

Most commonly, score-performance alignment is carried out automatically using software tools that compute the temporal alignment, which is then manually corrected if necessary. Multiple algorithms that approach this problem have been developed within the Music Information Research (MIR) literature (for an overview see Niedermayer, 2012), several of them being influenced by earlier algorithms used in speech processing.

[18] http://sourceforge.net/projects/wavesurfer
[19] http://www.sonicvisualiser.org

Practically all automatic score-performance alignment approaches can be formulated as the task of producing an association between the elements of two time-dependent sequences that represent the audio recording and the score, respectively. The differentiation between approaches comes from the following two factors:

- The **representation** of each sequence, and
- The **algorithm** that computes the optimal alignment between the two sequences.

The audio and score sequences can be represented in two ways: as a sequence of audio frames obtained by splitting a sound recording into consecutive chunks, each containing a fixed number of audio samples, or as a sequence of symbolic data. Niedermayer (2012) organizes score-performance alignment systems in three categories based on this: *Audio-to-audio* (A2A) systems, where the score is converted to sound using a synthesizer; *Audio-to-symbolic* (A2S) systems, where the audio recording is represented as a sequence of audio frames and is aligned to a symbolic representation of the score, and *Symbolic-to-symbolic* (S2S) systems, where note events are automatically transcribed from the recording and are aligned to a symbolic representation of the score.

In regard to the algorithm that computes the alignment, several approaches have been proposed, with two being the most ubiquitous: Hidden Markov models (HMMs) in combination with dynamic programming, and *Dynamic Time Warping* (DTW). HMM-based approaches represent note events from the score as a Markovian sequence of hidden states, and features extracted from frames of the audio recording as a sequence of observations on the hidden states; from there on, a dynamic programming routine (such as the *Viterbi* algorithm) is typically used to minimize the cost of the transitions (and thus maximize their probability) from state to state based on the observed audio feature frames (Cano et al., 1999; Orio and Déechelle, 2001). An overview of this approach can be seen in Figure 3.8.

Dynamic Time Warping is a technique used to align time series sequences to each other, first by computing the dissimilarity (or distance) between all frames of both sequences, and then calculating an alignment path that minimizes the cumulative distances of each pair of frames as that path progresses; a visual representation of this process can be seen in Figure 3.9. Score-performance alignment using DTW is therefore usually carried out by representing both the recording and the score as sequences of features extracted from their respective audio frames and calculating the distance between them in order to compute the alignment path (Dannenberg and Raphael, 2006; Dixon, 2005; Carabias-Orti et al., 2013).

Depending on the desired precision of the score-performance alignment, the ob-

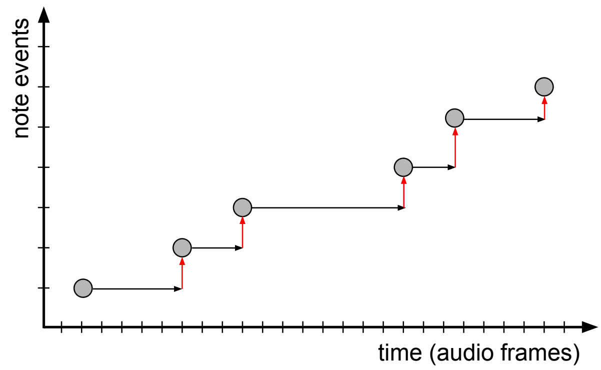


Figure 3.8: Visual representation of score-performance alignment using Hidden Markov Models. The score is represented as a sequence of note events/states (y-axis), while audio frames represent time (x-axis). Black arrows represent time spent in each state, while red arrows represent note transition events.

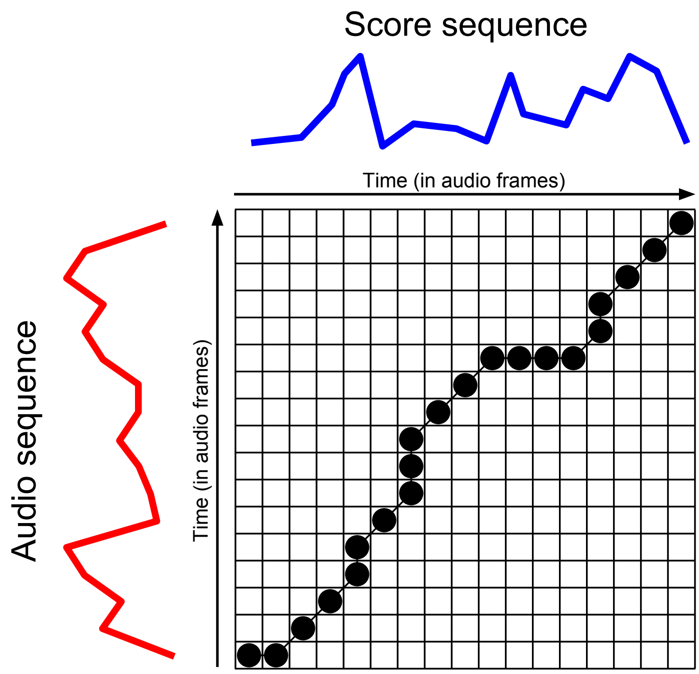


Figure 3.9: Visual representation of score-performance alignment using Dynamic Time Warping. Both the audio and the performance are represented as a sequence of audio frames. Black dots represent pairs of aligned audio frames along the alignment path.

tained note onset and offset times often require manual inspection and correction. For the score-performance alignment of our experimental recordings, we used a semi-automatic, conditional two-step approach that combines both DTW- and HMM-based methods. In the first step, each recording was aligned to its underlying score using an audio-to-audio alignment based on DTW; for simpler scores with few notes such as the exercises, this alignment was accurate enough to require minimal inspection and correction. For more complex scores such as the piece excerpts where more correction was necessary, the obtained note onset and offset times were re-aligned to the recorded performance utilizing both audio and instrumental gesture information (bow force, bow velocity) with an HMM-based implementation of a dynamic programming algorithm. In the rest of this section, we will provide some details on the DTW-based approach as well as a link to its open-sourced implementation; for a detailed account of the HMM-based alignment used for our experimental data see (Marchini, 2014).

3.3.1.2. Score-performance alignment based on Dynamic Time Warping

The DTW algorithm Basic sequence alignment using Dynamic Time Warping can be formulated as follows. Given two sequences $\{R = r_1, r_2, ..., r_m\}$ and $\{S = s_1, s_2, ..., s_n\}$, the goal is to find a minimum cost path $\{P = p_1, p_2, ..., p_k\}$. Each p_k is an ordered pair (i_k, j_k) such that $(i, j) \in P$ and the points r_i and s_j are aligned. The alignment is assessed with respect to a distance function $d_{R,S}(i, j)$, usually represented as an $m \times n$ distance matrix, which corresponds to the dissimilarity of each pair (p_i, r_j). The cost is 0 for a perfect match, and is otherwise positive. The total path cost $D(P)$ is the sum of the local costs along the path:

$$D(P) = \sum_{1}^{k} d_{R,S}(i, j) \tag{3.1}$$

In order to calculate the optimal path P, an $m \times n$ cost matrix $C_{R,S}$ is iteratively constructed using the above equation. Several constraints are placed on P, namely that the path must be:

- bounded by the ends of both sequences: $P_1 = 1, 1$ and $P_k = m, n$
- monotonic: $i_{k+1} \geqslant i_k$ and $j_{k+1} \geqslant j_k$
- continuous: $P_{k+1} - Pk \in (1, 0), (0, 1), (1, 1)$

Other local path constraints are also common, which alter the monotonicity and continuity constraints to allow increments of up to two or three steps in either direction and/or require a minimum of at least one step in each direction.

The minimum cost path is calculated using the following dynamic programming recursion:

$$D(i,j) = d(i,j) + min \begin{cases} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{cases} \tag{3.2}$$

DTW applied to audio In our implementation of audio-to-audio alignment, sequences R and S correspond to sequences of audio frames for the recording and synthesized score, respectively. Each element $r_i \in R$ and $s_j \in S$ corresponds to a feature vector representing either the *magnitude spectrum* of that audio frame, or a *chroma vector*, a feature vector with 12 elements containing the cumulative magnitudes for all frequencies corresponding to each of the twelve musical notes in a western classical scale (Bartsch and Wakefield, 2001). The choice of feature vector can be informed by two factors: (i) the length of the analyzed performance, and (ii) the acoustic fidelity of the synthesizer used to convert the score to audio:

- Chroma vectors, having only 12 values per frame, have lower memory requirements compared to the frequency magnitude spectrum (the size of which depends on the resolution of the frequency representation) and are thus optimal for longer pieces where the entire audio sequence must be loaded into memory.
- On the other hand, the low resolution of chroma vectors means that almost all information on the spectral envelope of the compared sounds is lost. For this reason, scores that are realistically synthesized with high quality soundbanks may justify the use of the entire spectrum as the feature vector in order to obtain a more detailed comparison between the recorded performance and the synthesized score.

Finally, euclidean distance or cosine distance can be chosen as the distance function $d_{R,S}(i,j)$ between the feature vectors for frames r_i and s_j.

On-line Time Warping variation on DTW Dixon (2005) proposed a variant of the DTW algorithm designed for real-time score-performance alignment called On-line Time Warping (OLTW). When the alignment is computed in real time, the sequence originating from the recorded performance R is not known and can therefore only be aligned up to the last received audio frame. In order to achieve this, a series of iterative DTW alignments are performed step-by-step as the R sequence progresses, using a subsequence of both R and S with a fixed search width. For each step of the DTW calculation the locally computed partial cost matrix is used to decide on the direction in which to proceed by advancing one frame further in the R sequence, the S sequence, or both. The decision is made by finding the direction in which lies the frame combination with the lowest cumulative cost thus far. Figure 3.10 shows a visual representation of the OLTW algorithm.

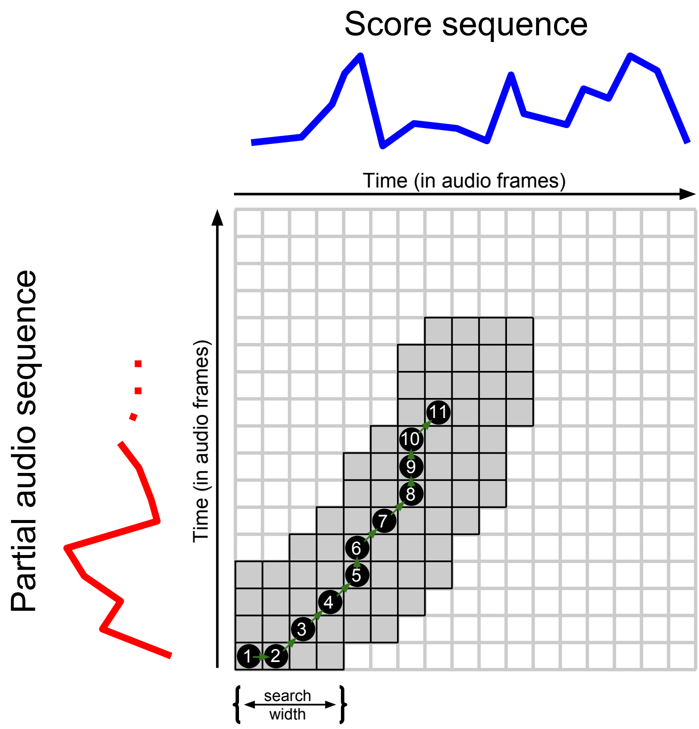


Figure 3.10: Visual representation of score-performance alignment using On-line Time Warping. Both the audio and the partially recorded audio of the performance are represented as a sequence of audio frames. Black dots represent pairs of aligned audio frames along the alignment path, which iteratively progresses in the direction shown by the green arrows.

For the case of our recordings, real-time alignment was not a concern as all recorded performances were aligned to their respective scores in a post-processing step; however, using the OLTW algorithm holds the advantage of having to compute only a subset of the distance matrix and cost matrix using sparse matrices, drastically reducing the memory requirements of the algorithm.

Software implementation of the OLTW algorithm We implemented a version of the OLTW algorithm based on pseudocode provided by Dixon (2005). The code is cross-platform, written in the *MATLAB*[20] programming environment, and can be freely accessed and downloaded as open source software under the AGPL license[21].

The aligner is provided as a software tool with a graphical user interface (GUI) that allows the user to set the following parameters of the alignment:

- The *frame size* (in samples) in which the recorded performance and synthesized score are split
- The *type of feature vector* (spectrum, chroma) extracted for each audio frame
- The *type of distance metric* used (euclidean, cosine)
- The *search width* of the OLTW algorithm

The user begins the alignment process by choosing an audio recording (in either .wav or .mp3 format) and a score in MIDI format, which is synthesized to audio using the *MuseScore*[22] notation software (on Windows) or the *FluidSynth*[23] software synthesizer (on OSX). The waveforms of both audio files are plotted on a panel of the GUI. Then, the user chooses the audio frame size and the type of feature vector to be extracted from each frame; the two vector sequences are then plotted on another panel for visual inspection of their overall similarity. Finally, the user chooses the distance metric to be used in the computation of the distance matrix, as well as the search width of the OLTW algorithm.

Once the alignment is computed, the user can then re-synthesize the now-aligned score and listen to it alongside the recorded performance in order to assess the accuracy of the alignment. The note onset and offset times are stored in a tab-separated text file along with the label of each aligned note, using a format compatible with the *Wavesurfer* annotation software (`ONSET_TIME OFFSET_TIME NOTE_LABEL`). The aligned score can also be saved in a format compatible with *Sonic Visualizer*. Figure 3.11 shows a screenshot of the aligner GUI where each of the above described components can be seen.

Aligning the experimental recordings All of the performances recorded during the main experiment were first aligned using the software tool described above.

[20] www.mathworks.com/products/matlab
[21] https://github.com/slowmountain/scorealigner
[22] https://musescore.org
[23] http://www.fluidsynth.org

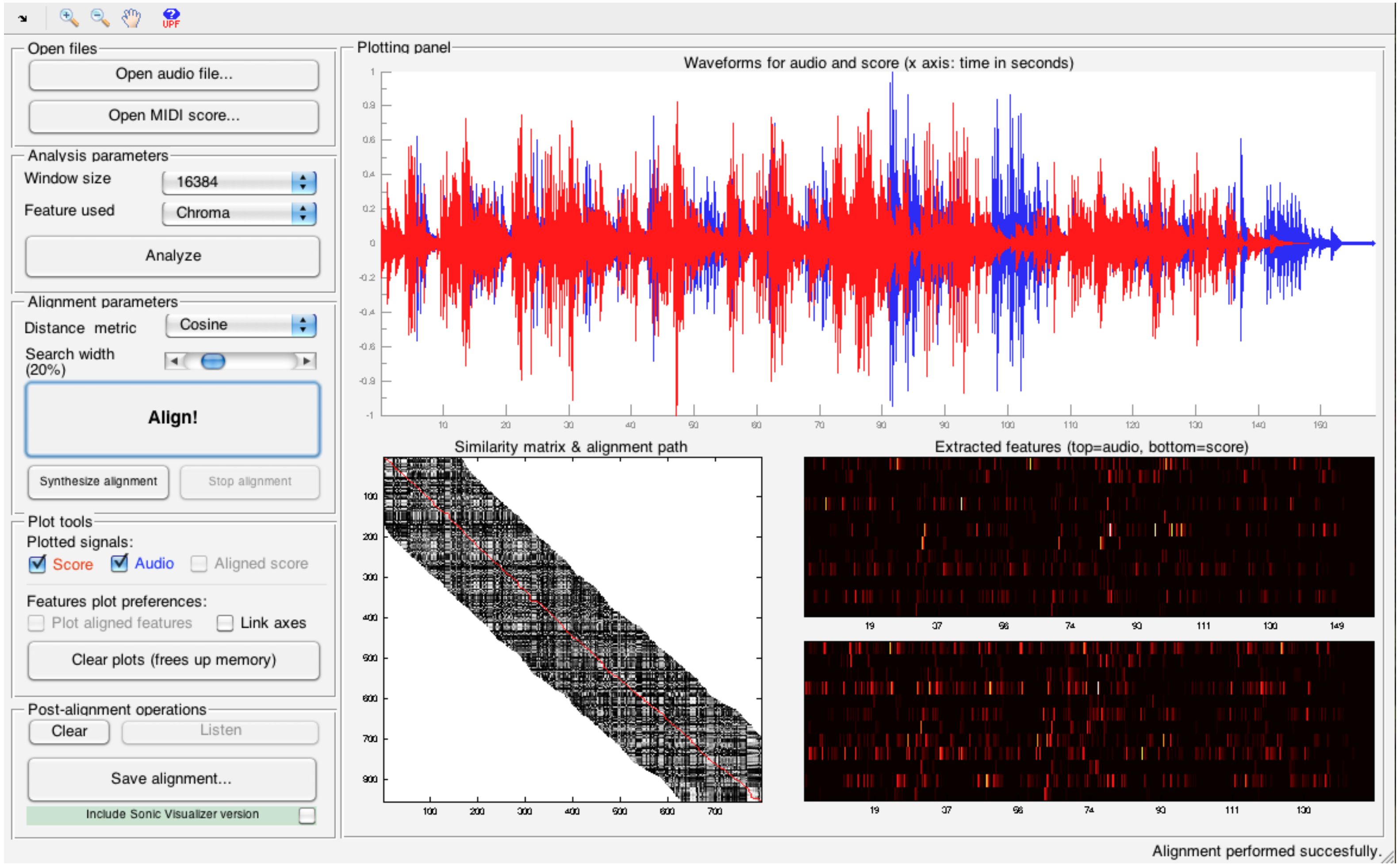


Figure 3.11: The graphical user interface of the implemented score alignment software. The waveforms of the performance recording (in blue) and the synthesized score (in red) can be seen in the top panel. The extracted feature vector sequences can be seen in the bottom right panels, and the iteratively computed distance matrix along with the final alignment path (in red) can be seen in the bottom left panel. The left part of the GUI contains the controls and alignment parameters for the tool.

Following the alignment, a manual evaluation of the alignment accuracy was carried out by listening to the synthesized aligned score and visually inspecting the detected note onset and offset times using the *WaveSurfer* software tool. In cases where the alignment was sufficiently accurate (i.e. note onset and offset errors in the range of ± 200 milliseconds), the alignment was manually corrected. In cases where the onset time errors were greater and the note density of the score was too high for efficient manual correction, the obtained alignment times were given as input to a more complex HMM-based score aligner that combined audio and instrumental gesture data; details on the algorithm itself and its implementation can be found in (Marchini, 2014).

3.3.2. Repository

All recordings in the main experiment have been uploaded to an online repository for multimodal data developed and maintained by the Music Technology Group of Universitat Pompeu Fabra, called *RepoVizz*[24] (Mayor et al., 2013).

RepoVizz represents multimodal recordings as *datapacks*, collections of time-synchronous multimodal data files (signals or annotations) organized in a tree structure that holds pointers to such files, associated metadata, text-based descriptions, and pointers to supplementary files. The tree structure is implemented by means of a custom XML file which is hierarchically formed by different types of nodes, with functions ranging from organizing data or holding textual descriptions to holding pointers to data files of different types. The different types of nodes inside a RepoVizz datapack structure can be seen in Figure 3.12, while a schematic representation of a RepoVizz datapack can be seen in Figure 3.13. Color coding is used for the different types of nodes.

[24]http://repovizz.upf.edu

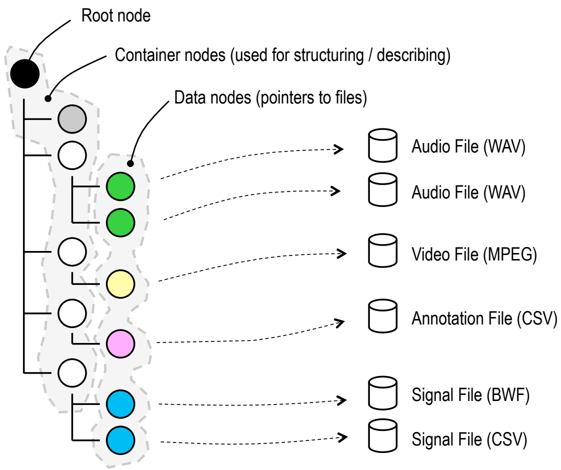


Figure 3.12: Schematic illustration of a RepoVizz datapack.

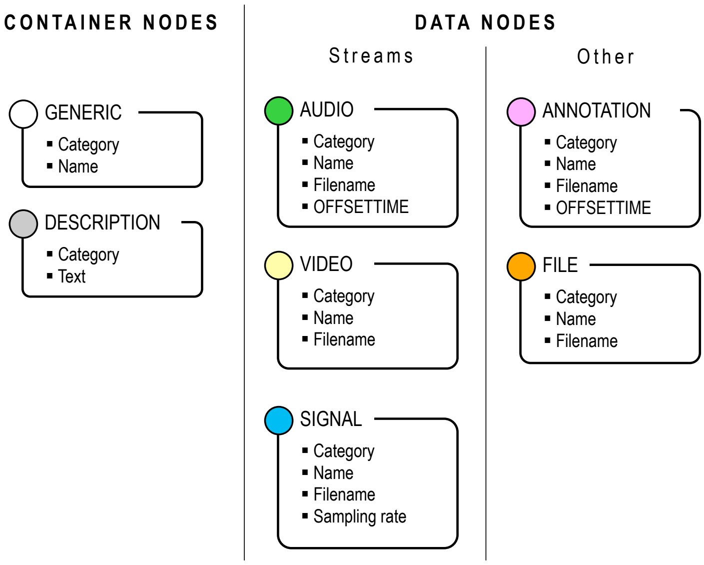


Figure 3.13: Overview of RepoVizz structure node types and corresponding attributes.

Besides storing and structuring multimodal data sets, RepoVizz also has a web-based client interface. Through it, users can browse datapacks, interactively visualize data streams and annotations and download segments of multiple data streams simultaneously. A screenshot of the web client interface showcasing the data visualization and exploration capabilities of RepoVizz for one of the experimental recordings (IP2) can be seen in Figure 3.14.

In addition to the uploaded data for each experimental recording, RepoVizz also computes audio features for every audio recording inside the datapack using the *Essentia* feature extraction library (Bogdanov et al., 2013).

3.3.3. QUARTET dataset guide

The QUARTET dataset can be accessed online at http://mtg.upf.edu/download/datasets/quartet-dataset, or directly from RepoVizz. For each recorded exercise and piece excerpt, five corresponding datapacks have been uploaded:

- Solo violin 1,
- Solo violin 2,
- Solo viola,
- Solo cello, and
- Ensemble.

A single exception to the above exists for the Intonation exercise (I1), where two ensemble datapacks exist - one without the visual annotations seen in Figure 3.1, and one with.

Each datapack has a short-form title that serves as an identifier, with each letter or number in the title denoting a characteristic of the datapack. For example, *D2S1* corresponds to *Dynamics exercise 2 Solo violin 1*, *D2S3* corresponds to *Dynamics exercise 2 Solo viola*, *D2E* corresponds to *Dynamics exercise 2 Ensemble*, while *DP2E* corresponds to *Dynamics piece excerpt 2 Ensemble*. A legend for the symbols used in the identifier of each datapack can be seen in Table 3.2.

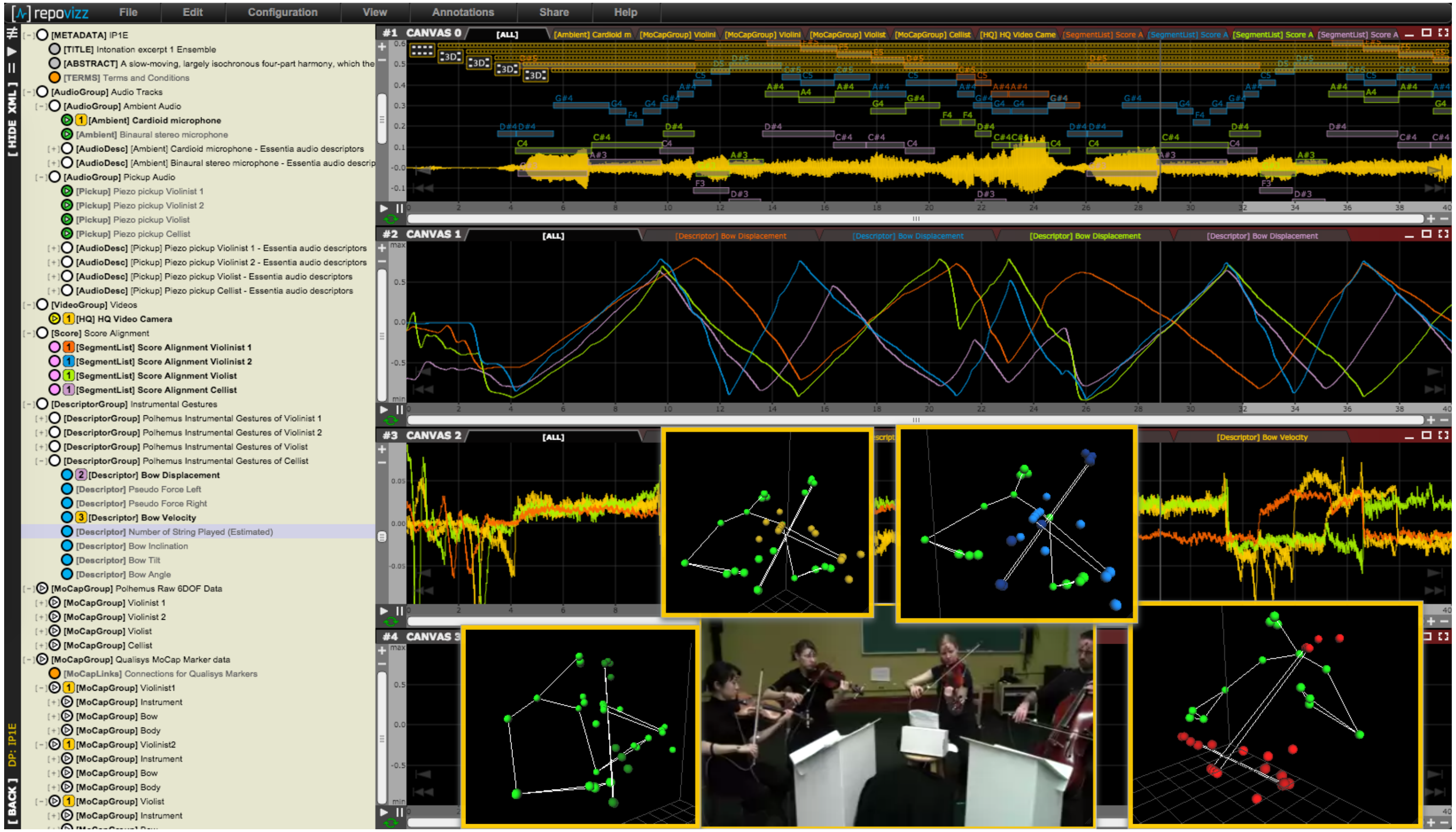


Figure 3.14: A screenshot of the datapack for recording IP2 accessed through RepoVizz, with several different multimodal data streams and annotations being visualized. On the left, the structure of the datapack is seen. On the top right canvas, the waveform of the cardioid microphone recording along with the aligned scores for all four parts are shown. On the two canvases below it, the bow displacement and bow velocity instrumental gesture features for all instruments are shown. Four floating windows can be seen on the bottom, each with a 3D visualization of the upper body and instrument motion capture data for each performer, while a fifth floating window with the video of the entire quartet is positioned between them.

Score identifier	Meaning
I, D, T, R, U, P	Intonation, Dynamics, Timbre, Rhythm, Unity of Execution, Phrasing
P (after the first letter)	Piece excerpt
S1, S2, S3, S4, E	Solo violin 1, Solo violin 2, Solo viola, Solo cello, Ensemble

Table 3.2: Legend for the short-form title identifier of each recording datapack.

Besides the short-form title, each shared datapack has a description (providing details on the datapack itself), an abstract (providing details on the recorded exercise/piece), and a list of keywords by which it is searchable.

Each recording contains a number of data files corresponding to the various data modalities captured in the experiment and detailed in Section 3.2.3. The files are organized inside the datapack structure in the following nodes:

- An `AudioGroup` node, containing the recorded audio from each of the utilized microphones (a cardioid microphone, a binaural microphone, and one bridge pickup microphone per performer)
    - Each audio recording is accompanied by an `AudioDesc` node, containing the audio features extracted using the *Essentia* library.
- A `VideoGroup` node, containing the recorded HQ video
- A `Score` node, containing the note onset and offset times for the audio of each instrument as they were obtained by the score-performance alignment
- A `DescriptorGroup` node, containing the instrumental gesture features for each instrument computed from data acquired using the wired motion capture system
- A `MoCapGroup` node, containing the raw 6DOF (position & orientation) data acquired using the *Polhemus* wired motion capture system
- A `MoCapGroup` node, containing the upper body marker data, instrument marker data, and bow marker data acquired using the *Qualisys* optical motion capture system

As a result of errors during the recording (especially in regard to the optical motion capture software and the video camera), some datapacks may not contain all of the above nodes. Audio recordings, score-performance alignment, instrumental gesture features and raw motion capture data from the wired system are present in all datapacks.

A complete index of all recordings with their respective durations can be found in Appendix C.

Chapter 4

INTERDEPENDENCE STUDIES

In this chapter we present the studies on the musical interdependence of string quartet ensembles that were carried out through the course of this dissertation, based on the experimental recordings that were described in Chapter 3.2. We begin by detailing the computational aspects of the methodology we applied to measure interdependence in terms of four dimensions of the performance (intonation, dynamics, timbre, timing).

We then present results from the application of this computational methodology on real performance data acquired in our experimental recordings, focusing on four exercise categories (intonation, dynamics, timbre and rhythm), each one corresponding to a performance dimension according to Table 3.1.

Finally, we present the results of an additional study that dealt with interdependence outside a performance analysis point of view, by assessing the capability of listeners to perceive the difference between ensemble string quartet performance and artificially synchronized solo performances.

4.1. Methodology

In Figure 1.2 of Chapter 1.5.2, we broke down the methodology followed in this dissertation into three steps: *performance capturing*, *performance description*, and *interdependence estimation*. Chapter 3.2 provided details on the performance capturing step, both in terms of design as well as implementation. In the next two subsections we will provide details on the performance description and interdependence estimation steps, respectively.

An overview of these two steps can be seen in Figure 4.1; starting from the ensemble and solo recordings of audio and instrumental gesture data, each recording is processed to extract a set of features per each performance dimension. Following feature extraction, features from the solo recordings are artificially synchronized to each other using the ensemble features as a reference. Then, we apply four different interdependence estimation methods (Pearson Correlation, Granger causality, Mutual Information and Nonlinear Coupling coefficient) to the ensemble and warped solo features, and compare the obtained values of interdependence between the solo and ensemble conditions using statistical methods.

4.1.1. Performance description

We analyze audio and instrumental gesture information from each recorded performance, to extract numerical features in the form of time series that describe the performance in terms of four dimensions: intonation, dynamics, timbre and timing. This subsection provides details on the feature extraction, organized by performance dimension.

Intonation refers to the accuracy of the produced pitch for a musical performance. We remind the reader that bowed string instruments are fretless and are therefore capable of producing continuous pitch; the musicians must constantly perform adjustments to their intonation in order to achieve harmonic consonance for the overall sound. When each performer has a different score, the achievement of consonance is further complicated by the choice of tuning temperament, as discussed in Chapter 2.1.1.1.

It must be pointed out that our goal is not to observe whether the performers are indeed achieving consonance or not, but to measure whether the adjustments to their intonation are influenced by other members of the ensemble. Given how each performer has their own part, a direct comparison of the produced pitch of each performer is not possible; an objective reference is needed in order to express intonation adjustments in a common scale across performers and different melodic lines. We utilize the notes in the score as "reference pitch", i.e. perfect non-adjusted intonation (according to the equal temperament tuning system), from which the deviations of the performer can be calculated.

Our pitch deviation feature is therefore defined as follows. For every recording we estimate pitch from the bridge pickup audio signal using the YIN algorithm (De Cheveigné and Kawahara, 2002), which divides the signal into frames of 1470 samples with a hop size of 32 samples, and yields a pitch value for each frame expressed in a linear scale of octaves (with 0 being 440 Hz, 1 being 880 Hz etc.).

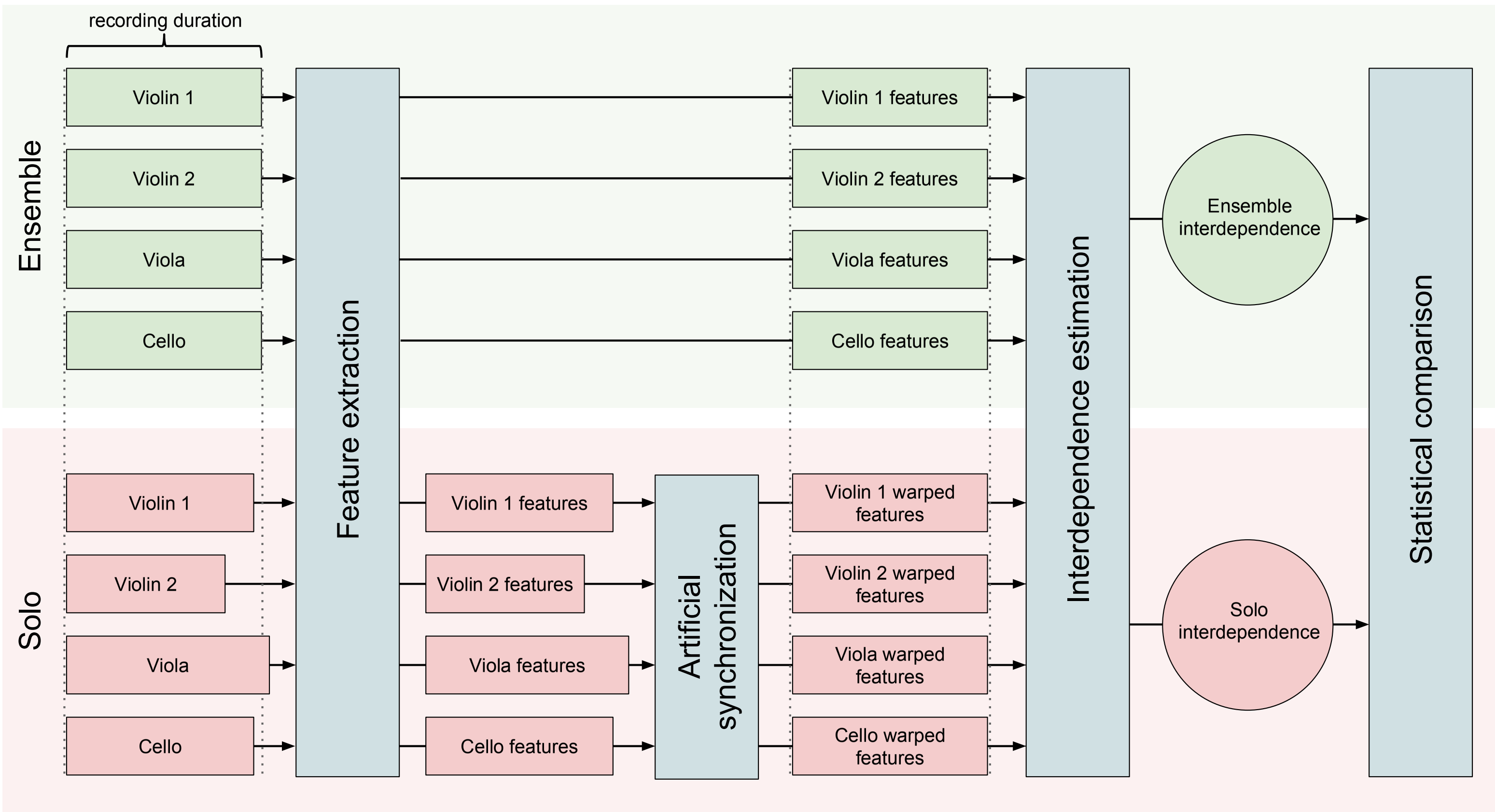


Figure 4.1: Schematic representation of the Performance Description and Interdependence Estimation steps of our computational methodology. Green blocks and red blocks represent ensemble and solo data, respectively, while blue blocks represent processing steps.

Let us define $p(t)$ as the estimated pitch for the t-th frame of a recording. For the i-th note in in the score-performance alignment, let t_0^i and t_1^i be the note onset and note offset times, respectively, and m^i be the MIDI pitch value of that note. We compute the reference pitch $s(t)$ as a note-by-note piecewise constant time series such that $s(t) = \dfrac{m^i - 69}{12}, \quad t \in [t_0^i, t_1^i]$. The number 69 corresponds to the MIDI pitch value of note A4 (440 Hz), which corresponds to a reference pitch value of 0. By computing the difference between $p(t)$ and $s(t)$, we obtain the pitch deviation $\hat{p}(t)$:

$$\hat{p}(t) = p(t) - s(t) \tag{4.1}$$

.

An example of the pitch deviation feature for the D1 exercise can be seen in Figure 4.2.

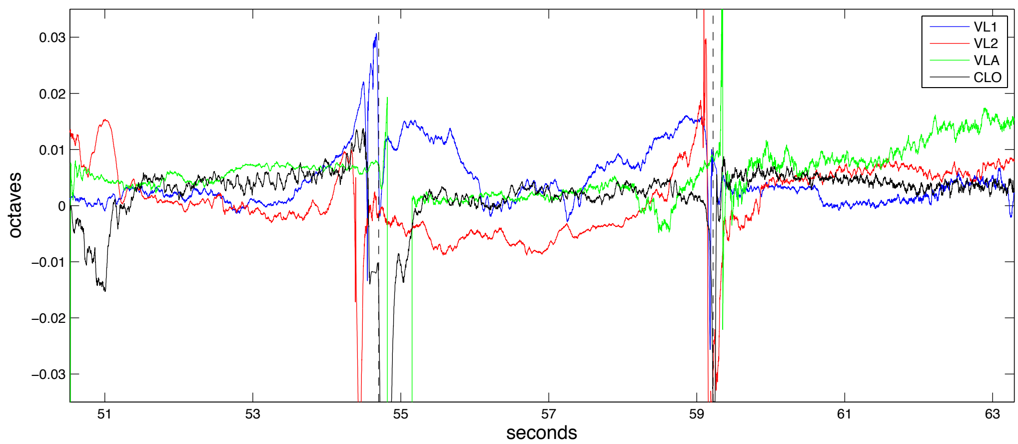


Figure 4.2: Excerpt of the pitch deviation feature from the ensemble recording of exercise D1, for all instruments. Dashed vertical lines represent note boundaries. VL1 corresponds to violin 1, VL2 to violin 2, VLA to viola and CLO to cello.

Dynamics refer to the varying levels of sound intensity in a musical performance. For every recording, we estimate sound intensity (in dB) as follows. The bridge pickup audio signal is divided into frames of N = 1470 samples with a hop size of 32 samples as in the case of the pitch deviation feature. Let $[x_1(t), \ldots, x_N(t)]$ be the samples of the t-th frame. For each frame we compute the the sound intensity $d(t)$ in a logarithmic scale of dB:

$$d(t) = 20 \log_{10} \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x_n(t))^2} \tag{4.2}$$

The logarithmic scaling is performed in order to obtain an estimation of loudness that is closer to human auditory perception, as per Fechner's law (Dehaene, 2003).

Typically, musical scores include both instantaneous as well as gradual changes in dynamics. When the scores of more than one performer feature the same dynamics change at the same time, this can give the impression of coordinated action regardless of whether such coordination actually exists. In order to remove such a bias from our data, we detrend d on a note-by-note basis as follows. For each note i in the score-performance alignment of the recording, we obtain the note onset time t_0^i and note offset time t_1^i. We then use linear regression on $d(t)$ to obtain a straight-line fit $l(t)$, and construct the detrended feature $\hat{d}(t)$:

$$\hat{d}(t) = d(t) - l(t), \quad t \in [t_0^i, t_1^i] \tag{4.3}$$

This way, note-to-note changes in dynamics are greatly reduced, making temporal fluctuations of dynamics within the boundaries of each note much more prevalent. An example of the detrended sound intensity feature for the D1 exercise can be seen in Figure 4.3.

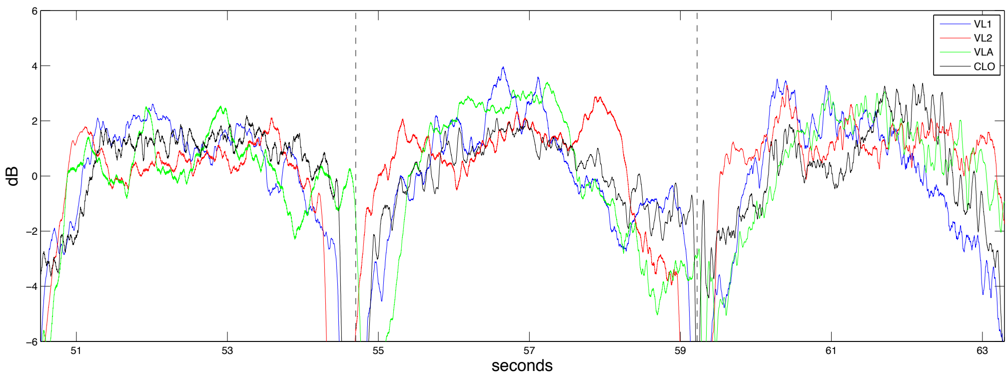


Figure 4.3: Excerpt of the detrended sound intensity feature from the ensemble recording of exercise D1, for all instruments. Dashed vertical lines represent note boundaries. VL1 corresponds to violin 1, VL2 to violin 2, VLA to viola and CLO to cello.

Timbre refers to the character or quality of the produced sound distinct from its pitch and intensity. As a dimension of musical performance it is challenging to fully describe, both from a semantic as well as a computational point of view. In our analysis, we opted to quantify the timbre characteristics of the performance using two different types of data: instrumental (sound-producing) gesture data as well as audio data.

Our gesture-based timbre feature is the *bow-bridge distance*, the distance (in centimeters) between the bridge and the point of contact between the bow and the string that is being excited. Composers and performers use the *sul tasto* or *sul ponticello* (on the bridge and on the fingerboard, respectively) terms to describe bow-bridge distance in written music and manipulate the "brightness" of the produced sound, by affecting the distribution of energy in the sound spectrum between higher and lower harmonics. Bow-bridge distance manipulation is also the main focus of the timbre exercise from our experimental recordings. An example of the bow-bridge distance feature for the D1 exercise can be seen in Figure 4.4.

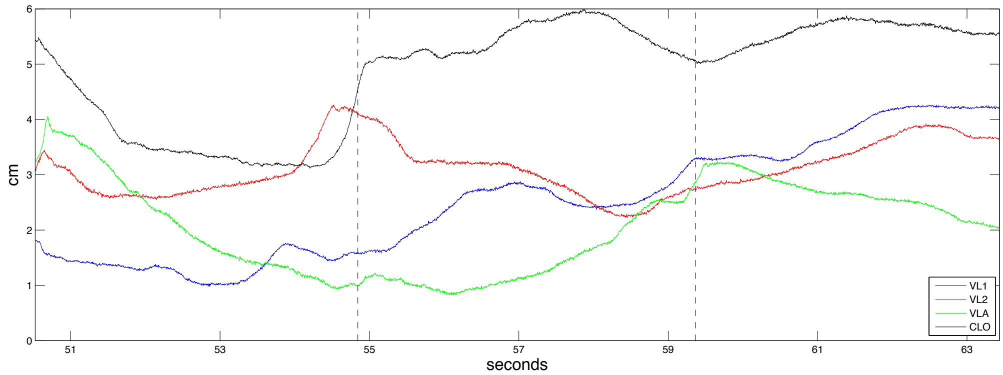


Figure 4.4: Excerpt of the bow-bridge distance feature from the ensemble recording of exercise D1, for all instruments. Dashed vertical lines represent note boundaries. VL1 corresponds to violin 1, VL2 to violin 2, VLA to viola and CLO to cello.

Audio-based timbre features are typically extracted from a frequency domain representation: examples of the most prevalent features include the *Spectral Centroid* and the *Mel-Frequency Cepstral coefficients* (MFCCs). However, separation between timbre and other dimensions of the performance (dynamics and pitch) is difficult to achieve in a spectral representation; recent work by Peeters et al. (2011) investigated the redundancy among spectral features using correlation analysis and hierarchical clustering, distinguishing ten classes of features that were largely independent from a statistical point of view. Of the available features based on their feature extraction toolbox, we chose to use the spectral crest feature: a descriptor of the noisiness/harmonicity of the sound that was shown to be relatively independent from other dynamics- or pitch-related features.

Spectral crest is calculated using a publicly available feature extraction toolbox[1] by Peeters et al. (2011) as follows: first, the audio signal is divided into frames of 4096 samples with a hop size of 184 samples (in order to obtain the same sampling rate of the instrumental gesture features). For the t-th frame of the audio signal, the power spectrum is obtained using the Short-time Fourier Transform (STFT), yielding K magnitude values $a_k(t)$ (where K is the number of STFT bins). The spectral crest $c(t)$ for that frame is obtained as the ratio between the maximum value and arithmetical mean of the power spectrum:

$$c(t) = \frac{\max a_k(t)}{\frac{1}{K}\sum_{k=1}^{K} a_k(t)} \tag{4.4}$$

An example of the spectral crest feature for the D1 exercise can be seen in Figure 4.5; as a ratio, it is dimensionless and therefore has no unit.

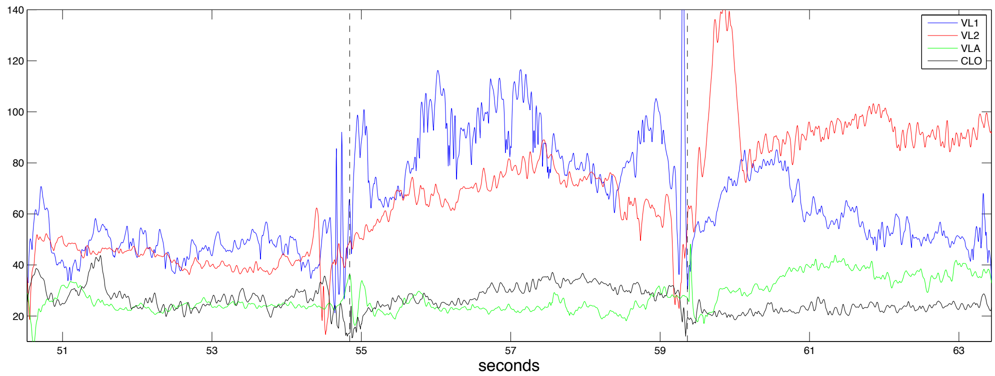


Figure 4.5: Excerpt of the spectral crest feature from the ensemble recording of exercise D1, for all instruments. Dashed vertical lines represent note boundaries. VL1 corresponds to violin 1, VL2 to violin 2, VLA to viola and CLO to cello.

Timing in music performance describes how the onset times and durations of notes deviate from a "mechanical" performance with absolutely constant tempo and strict adherence to the ratios between note values in the score. On a note-by-note basis, timing behavior can be described using the score-performance alignment by measuring the difference between the temporal position of the note onsets-offsets as performed by the musician, and the temporal position of the note onsets-offsets as the score defines them. We choose to study timing not on a note-by-note basis

[1]The Timbre toolbox: http://www.cirmmt.org/l/research-tools/timbretoolbox

but over the entire performance, by computing fluctuations of the pace or *tempo* at which the composition is performed throughout its duration.

In the context of musical feature extraction, tempo is typically represented with so-called *tempo curves*, measured in beats-per-minute (BPM). For the i-th note in a recording, we obtain the note onset time t_0^i and note offset time t_1^i from the score-performance alignment. We additionally obtain the reference note onset time $t_0^i|_r$ and reference note offset time $t_1^i|_r$ from the original score of the performance prior to the alignment, as well as the suggested tempo of the original score in BPM B_s. The tempo curve $B(i)$ is computed on a note-to-note basis as follows:

$$B(i) = B_s \frac{t_1^i - t_0^i}{t_1^i|_r - t_0^i|_r} \tag{4.5}$$

After the initial computation, the tempo curve is smoothed at the bar level using a moving average to derive an overall tempo behavior; this way we can focus on long-term tempo fluctuations (similar to the *period correction* mechanism discussed in Chapter 2.1.2.1) rather than short-term asynchronies. An example of the Tempo curve feature for the D1 exercise can be seen in Figure 4.6.

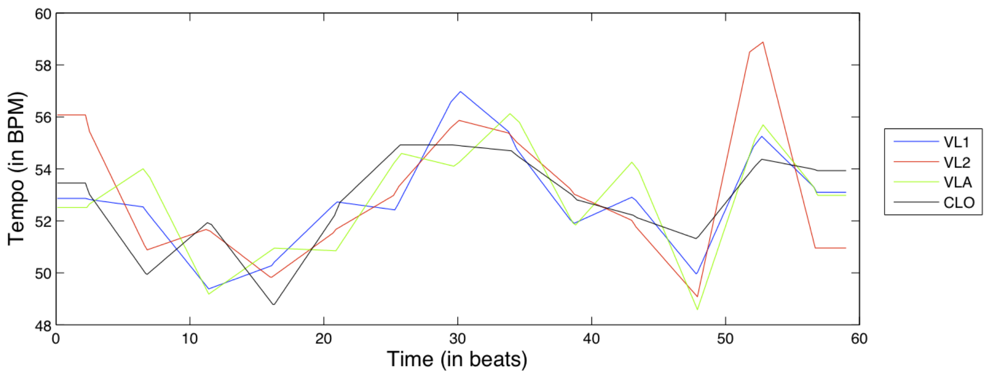


Figure 4.6: Tempo curve feature from the ensemble recording of exercise D1, for all instruments. VL1 corresponds to Violin 1, VL2 to Violin 2, VLA to Viola and CLO to Cello.

4.1.1.1. Artificial synchronization of solo features

In order to estimate interdependence between the four extracted features (one per performer) in a given performance dimension, it is necessary to ensure that they correspond to the same temporal axis in reference to the score. For the case of the ensemble recordings, this is guaranteed since the musicians performed simultaneously. However, for the case of the solo recordings, each musician performed alone; even though four bars of a metronome signal were provided prior to the beginning of the recording, given time the recordings begin drifting apart up to the point where the same sample index no longer corresponds to the same part of the score. For this reason, it was necessary to artificially synchronize extracted features from the solo recordings using a common reference.

For the intonation/dynamics/timbre dimensions, our objective was to study musical interdependence excluding timing information as much as possible; for this reason it is desirable to use features from the ensemble recordings as the common synchronization reference to which features from the solo recordings are artificially synchronized, so that the only difference between ensemble and solo are in terms of the studied performance dimension. For the case of the timing performance dimension, artificial synchronization of the solo tempo curve features is not necessary since they as well as the ensemble tempo curve features are expressed in beat positions of the score rather than absolute time.

The artificial synchronization of the solo features is carried out in a note-by-note fashion, using piecewise linear interpolation to resample the time series features between note onsets. Let $v(t)$ be the extracted feature time series from a solo recording, i a note from the score, $t_0^i|_s$ and $t_1^i|_s$ the onset and offset times of that note according to the score-performance alignment of the solo recording, and finally $t_0^i|_e$ and $t_1^i|_e$ the onset and offset times of that note according to the score-performance alignment of the ensemble recording. For each i, the resampled values of the artificially synchronized feature vector $w(t)$ are obtained as follows:

$$w(t_e) = f(t_s, v(t_s), t_e), \quad t_0^i|_e \leq t_e \leq t_1^i|_e \quad and \quad t_0^i|_s \leq t_s \leq t_1^i|_s \tag{4.6}$$

where f is a piecewise linear interpolation function that resamples the values $v(t)$ at time indices $t_s \in [t_0^i|_s, t_1^i|_s]$ to match time indices $t_e \in [t_0^i|_e, t_1^i|_e]$.

The above routine was directly applied to the spectral crest feature c and the bow-bridge distance feature; for the case of the detrended dynamics feature $\hat{d}$ and pitch deviation feature $\hat{p}$, the artificial synchronization routine was applied to the original features d and p, respectively.

A comparison between ensemble features and artificially synchronized solo features for the violin 1 recording of exercise D1 can be seen in Figure 4.7, while a comparison between artificially synchronized solo features and original solo features can can be seen in Figure 4.8.

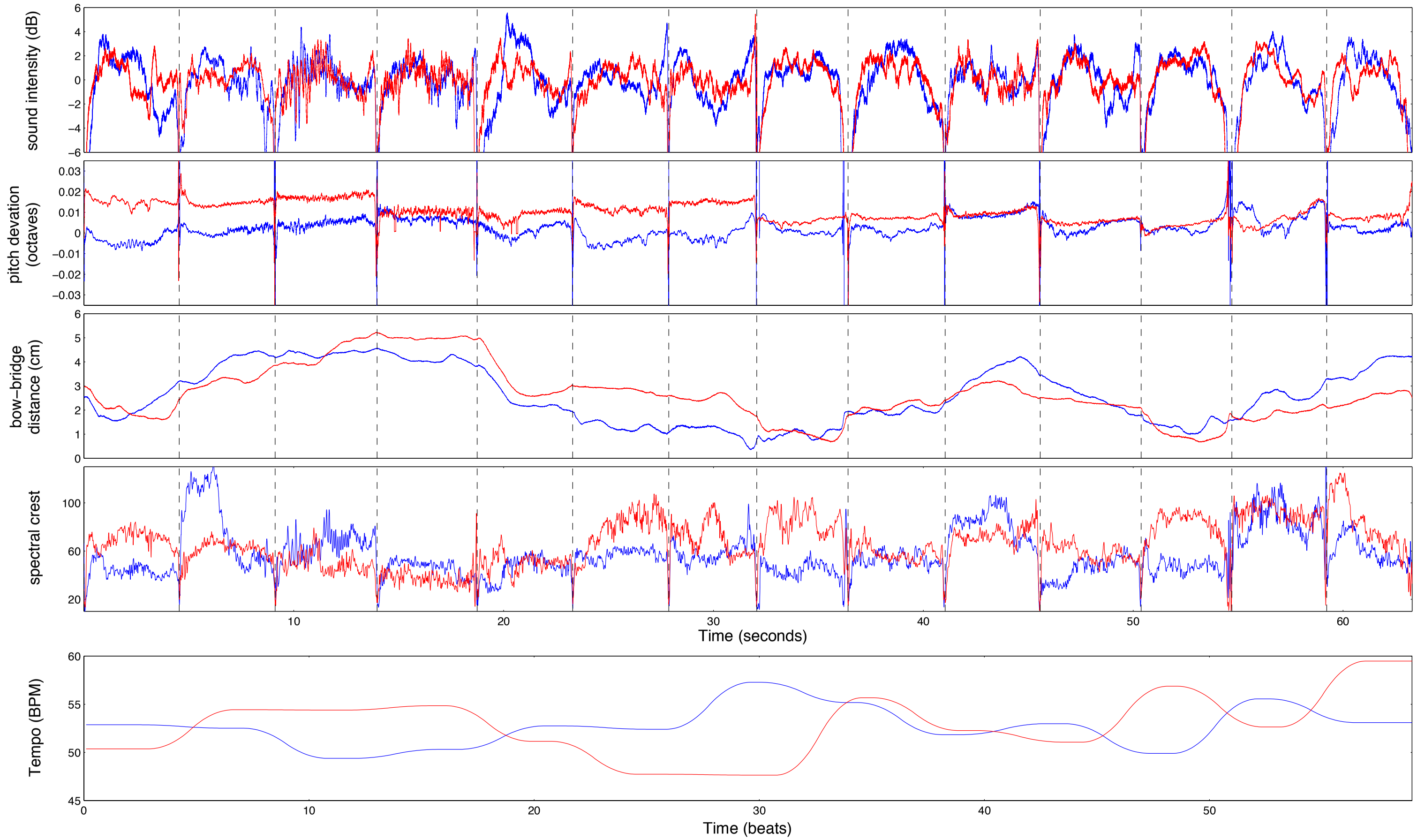


Figure 4.7: Comparison of ensemble features (in blue) and artificially synchronized solo features (in red) of violin 1, for exercise D1. Dashed vertical lines represent note boundaries.

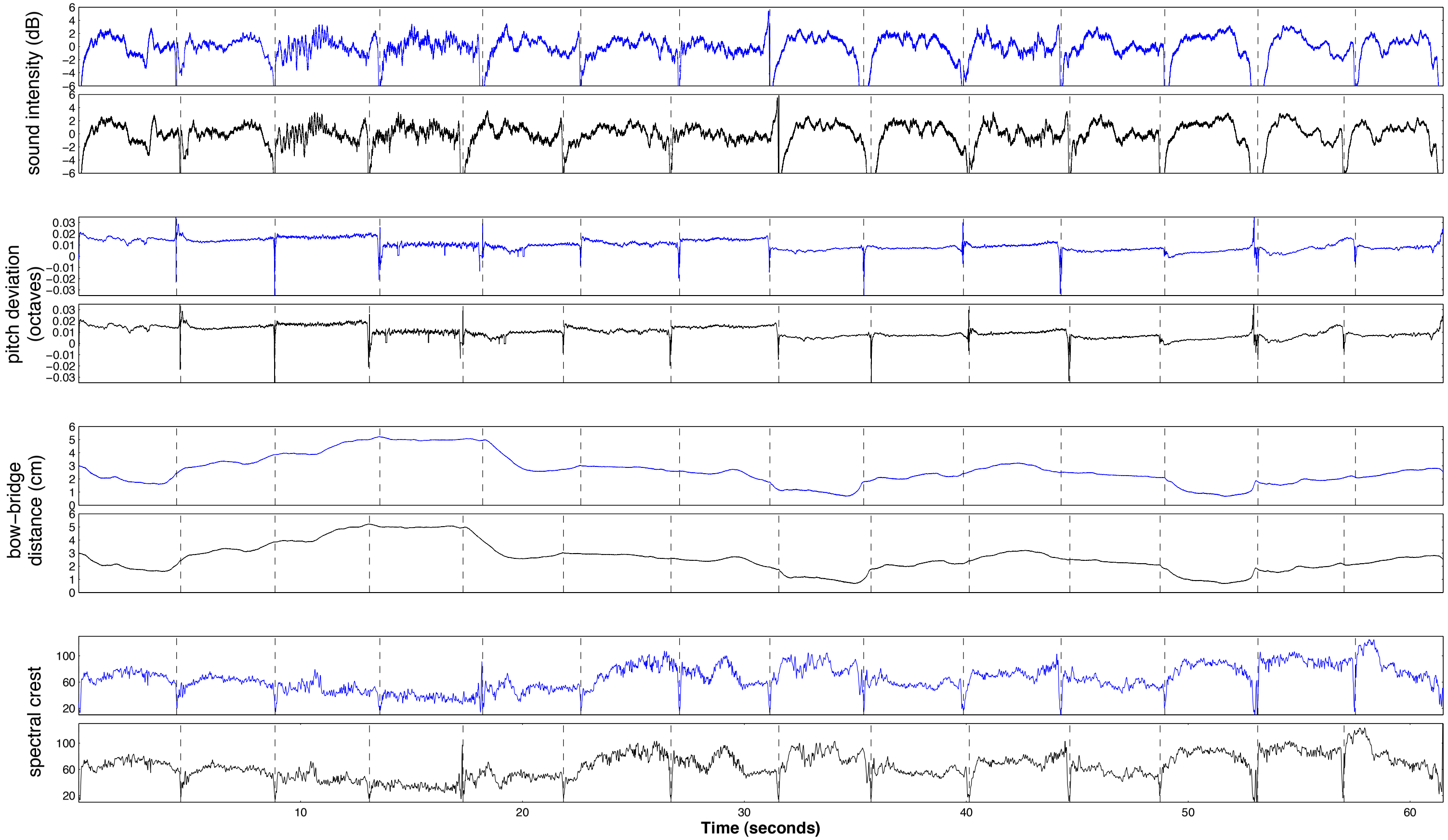


Figure 4.8: Comparison of artificially synchronized features (in blue) and original (in black) features for the solo recording of violin 1, for exercise D1. Dashed vertical lines represent note boundaries.

4.1.2. Interdependence estimation

Both for the ensemble and artificially synchronized solo features of a recording, interdependence between performers is estimated using four different methods: Pearson correlation, Granger causality, Mutual Information, and Nonlinear Coupling coefficient - for a presentation of each method see Chapter 2.3. Interdependence among the four musicians of the string quartet is estimated separately for each feature, once for the ensemble recording and once for the solo recording. The rest of this section provides details on how interdependence is estimated according to the applied method and studied performance dimension.

4.1.2.1. Overview

Given a segment s of a recording and a type of feature f, we obtain four time series f^s_x, $x \in [1, 4]$ for violin 1, violin 2, viola and cello, respectively. Interdependence between the four features is estimated for each of the four methods in a different manner, depending primarily on the directionality of the interdependence estimation method; below we detail the manner in which interdependence is estimated using each method, as well as the manner in which we obtain an estimation of the overall interdependence sustained by the entire string quartet for the analyzed recording segment.

Non-directional interdependence measures
For the case of Pearson correlation, the correlation coefficient ρ is estimated in a pairwise fashion to yield six symmetric assessments $\rho_{f^s_x f^s_y}$ where $x, y \in [1, 4]$ and $x < y$. The obtained values are additionally normalized using the Fisher z-transformation. In order to get an estimation of the overall interdependence, we average the six values across pairs to obtain $\overline{\rho}_s$.

Mutual Information between the four features is computed similar to Pearson correlation: first we obtain six pairwise assessments of Mutual Information $MI_{f^s_x f^s_y}$. Finally, we get an estimation of the overall interdependence by averaging across performer pairs to obtain $\overline{MI}_s$.

Directional interdependence measures
Interdependence estimation using Granger causality yields 12 asymmetric pairwise causality assessments (one per ordered musician pair) $G_{f^s_x, f^s_y}$, where $x, y \in [1, 4]$ and $x \neq y$. Additionally, Granger causality computes the total causal density G^d_s of the analyzed segment s, a bounded value between 0 and 1 with 0 signifying a

complete lack of causal interactivity; we use this value as an estimation of the total amount of causal interactivity sustained by the entire quartet, avoiding the need to average the individual pairwise assessments.

Similar to the case of Granger causality, the Nonlinear coupling yields 12 asymmetric pairwise calculations of coupling $L_{f_x^s, f_y^s}$. We currently have no method of measuring the overall coupling of the string quartet; the average value of the set of 12 coupling values $\overline{L}_s$ is used as estimation instead.

Parameter selection

Depending on the method used, it is possible that additional parameters are necessary to be defined prior to the computation. For the case of Granger causality, the order of the multivariate regression model applied by the method must be chosen; we used the Bayesian Information Criterion in order to automatically select the model order, as it was estimated from each of the features.

For the computation of the Nonlinear Coupling coefficient, there are four parameters (see Chicharro and Andrzejak, 2009) that must be given as an input to the algorithm:

- *embedding dimension (m)*, the number of past values to be included for the state space reconstruction
- *time delay parameter (τ)*, the time delay in samples between each of the past values used
- *number of nearest neighbors (k)*, and
- *Theiler window (W)*, the threshold used to exclude temporally close points from being selected as nearest neighbors.

Experimenting with the values of these parameters for the Nonlinear Coupling coefficient computation, it became evident that the parameters with the greatest impact on the outcome of the algorithm were the embedding dimension m and the time delay τ; the number of nearest neighbors was set to $k = 3$, and the Theiler window to $W = 2\tau$. We experimented with a wide range of values for both m and τ ($m \in [3, 10]$ and $\tau \in [2, 10]$), and found that the differences between ensemble and solo recordings were similar irrespective of the m and τ values (an example of this observation for the Dynamics dimension of the D1 excercise can be seen in Figure 4.9. The final parameters used were $m = 10$ and $\tau = 5$.

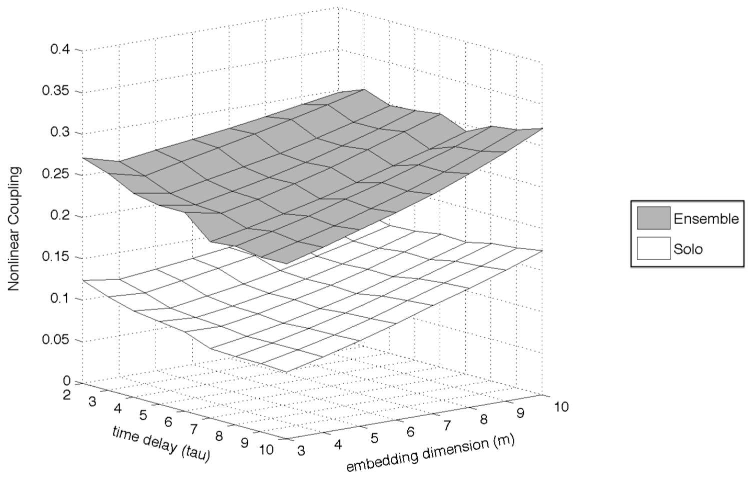


Figure 4.9: Example of the effect of parameter choice (embedding dimension and time delay) on the estimation of the Nonlinear Coupling coefficient, for the dynamics dimension of exercise D1.

4.1.2.2. Intonation, Dynamics and Timbre

For the intonation, dynamics, and timbre dimensions, we sequentially compute interdependence on the time series features by splitting them into consecutive non-overlapping windows. There are three reasons behind this decision: first, as we are studying features that change with time it is natural that the amount of interdependence will also vary - something which could potentially reveal the role of the musical score in the performance, as the collaborative task that must be jointly carried out by the musicians. Second, by windowing the time series we can reduce the non-stationarity in our data, thus making the interdependence measures more reliable. Finally, we can deal with smaller amounts of data at a time, which removes the need to downsample our signals in order to cope with memory requirements.

For these three performance dimensions, each segment s (as described in the previous subsection) corresponds to the samples of the time series between the boundaries of an analysis window. The window size is calculated in beats as obtained by the score-performance alignment; beat positions are converted to sample indices in order to calculate the beginning and end of each analysis window.

For each window segment s, overall interdependence is computed using the four methods as described in the previous subsection, yielding four values: $\overline{\rho}_s$, $\overline{MI}_s$,

G_s^d and $\overline{L}_s$. Finally, interdependence across the duration of a recording with N analysis windows yields four sequences of estimated interdependence: $[\overline{\rho}_1, \ldots, \overline{\rho}_N]$, $[\overline{MI}_1, \ldots, \overline{MI}_N]$, $[G_1^d, \ldots, G_N^d]$ and $[\overline{L}_1, \ldots, \overline{L}_N]$.

All of the recorded exercises have a bar length of four beats, with the exception of the intonation (I1) exercise which has a bar length of six beats; we chose a universal window size of eight beats as a compromise (in all exercises except I1, 8 beats correspond to two bars, while in I1 they correspond to 1.5 bar). Subsection 4.2.2 provides further information on how the window size affects the interdependence estimation.

4.1.2.3. Timing

The tempo curve feature used in the timing dimension is sampled at much large intervals as compared to the other features, and therefore consists of significantly fewer samples; for this reason, it was not feasible to divide the features in windows. The four interdependence values $\overline{\rho}$, $\overline{MI}$, G^d and $\overline{L}$ are therefore computed on tempo curves that represent the entire recording.

4.2. Results

In this section we report the results from our computational analyses of interdependence on the exercise recordings carried out as part of our main experiment.

Throughout the duration of our work, we published several studies on computational analyses of interdependence. A first study (Papiotis et al., 2011) focused solely on the performance dimension of intonation, and applied the Nonlinear Coupling coefficient on data acquired from the pilot experiment discussed in Chapter 3.1. A second study (Papiotis et al., 2012a) focused both on intonation as well as dynamics, and applied all four interdependence measures on a subset of the exercise recordings carried out during our main experiment. A third, higher-level study (Papiotis et al., 2013b) applied Mutual Information on all exercise data for all performance dimensions, and compared interdependence results across performance dimensions and exercises. Finally, an extended analysis using all four performance dimensions on the exercise data was carried out in a fourth study (Papiotis et al., 2014).

The methodology followed in each study improved on the foundations of the previous one by elaborating on the studied materials, performance dimensions considered, and interdependence methods applied. In this section we report results in an integrated fashion, minimizing the overlap between the studies by reporting results obtained from the most up-to-date iteration of our methodology as it was detailed in the previous section. For specific details on each of the individual studies, we refer the reader to the original publications where they were presented.

In the rest of this section we focus on the four exercise categories (intonation, dynamics, timbre, and rhythm) that are directly related to the four performance dimensions, as discussed in Chapter 3.2.1.2. We carry out an in-depth comparison of solo versus ensemble interdependence for all four interdependence methods using statistical methods. Additionally, we study the way in which ensemble interdependence varies across all of the studied exercises for each performance dimension individually, using the solo recordings as a baseline reference of interdependence.

4.2.1. Statistical analysis of interdependence in solo vs. ensemble recordings

4.2.1.1. Overview

We analyze six recorded exercises, each one relating to a specific performance dimension: exercise I1 for the intonation dimension, exercises D1 and D2 for the dynamics dimension, exercise T1 for the timbre dimension, and exercises R1 and R2 for the timing dimension. For each exercise, we estimate interdependence for the ensemble recording as well as for the artificially synchronized solo recordings. A comparison between interdependence results for the ensemble and solo recordings is then carried out to assess whether there is a significant difference in terms of interdependence between the two conditions.

We remind the reader that for the intonation, dynamics and timbre dimensions, interdependence is estimated in a windowed fashion, yielding a series of overall interdependence values along the duration of each studied exercise; a window size of 8 beats was chosen for all studied exercises, as discussed in Section 4.1.2.2. For each exercise where windowing was applied, we report the mean μ and standard deviation σ of overall interdependence averaged through all analysis windows, along with the number of interdependence values N corresponding to the number of windows.

We additionally report the results of a repeated measures ANOVA test on the results of all analysis windows, in order to assess whether there is a significant difference in average interdependence values between the ensemble and solo conditions. All ANOVA results reported in this section were additionally confirmed using a non-parametric (Wilcoxon rank-sum) test, which yielded similar results.

For the timing dimension, a windowed analysis of interdependence was not carried out, as the number of samples in the tempo curve features was too small to allow it. We thus simply report the results obtained from each interdependence method.

4.2.1.2. Results

Intonation

The four interdependence methods were applied on the pitch deviation feature as it was extracted from the ensemble and solo recordings of the I1 exercise. We remind the reader that for I1 two separate ensemble recordings were carried out, a first recording without annotations, and a subsequent recording with specific annotations regarding intonation behavior. Table 4.1 contains the averaged interdependence values for each method; a number of 17 interdependence values was obtained from the windowed analysis.

Method	**Interdependence strength, Intonation (I1)**					
	solo		**ensemble 1**		**ensemble 2**	
	μ	σ	μ	σ	μ	σ
Pearson correlation	0.001	0.321	-0.039	0.512	-0.032	0.564
Granger causal density	0.008	0.008	0.032	0.030	0.032	0.028
Mutual information	0.277	0.103	**0.415***	0.147	**0.464***	0.211
Nonlinear coupling coefficient	0.201	0.068	**0.263***	0.072	**0.287***	0.098

Table 4.1: Mean (μ) and standard deviation (σ, $N = 17$) of interdependence strength for the Intonation dimension per experimental condition (ensemble 1 - without annotations, ensemble 2 - with annotations). Values with an asterisk denote a statistically significant difference between ensemble and solo.

On initial inspection, it can be observed that the Linear Correlation method is not able to detect any significant differences between the ensemble and solo recordings of I1. Moreover, marginally negative correlation values between the pitch deviations of the musicians are observed in all three recordings.

Quite the opposite result can be observed for each of the three remaining interdependence methods. All of them estimate a higher amount of interdependence for the two ensemble recordings in comparison to the solo recording, while the ensemble-with-annotations recording demonstrates a slightly higher amount of interdependence in comparison to the ensemble-without-annotations recording.

The ANOVA analysis showed that both Mutual Information ($F(1, 16) = 18.324, p = 0.001, \eta^2 = 0.71$) as well as the Nonlinear Coupling Coefficient ($F(1, 16) =$

$16.085, p = 0.001, \eta^2 = 0.72$) could successfully separate the interdependence means of each ensemble condition from the solo condition. One the other hand, Granger causality did not provide significant separation between ensemble and solo. None of the interdependence methods found a statistically significant difference between ensemble-with-annotations and ensemble-without-annotations.

Dynamics

The four interdependence methods were applied on the detrended sound intensity feature as it was extracted from the ensemble and solo recordings of exercises D1 and D2. Table 4.2 contains the averaged interdependence values for each method; a number of 7 (D1) and 6 (D2) interdependence values were obtained from the windowed analysis.

Method	Interdependence strength, Dynamics (D1)				Interdependence strength, Dynamics (D2)			
	solo		ensemble		solo		ensemble	
	μ	σ	μ	σ	μ	σ	μ	σ
Pearson correlation	0.154	0.232	0.343	0.375	0.279	0.349	0.368	0.451
Granger causal density	0.002	0.001	0.004	0.004	0.003	0.001	0.008	0.002
Mutual information	0.306	0.067	**0.494***	0.092	0.449	0.109	**0.688***	0.182
Nonlinear coupling coefficient	0.152	0.042	**0.262***	0.052	0.256	0.046	**0.372***	0.078

Table 4.2: Mean (μ) and standard deviation (σ, $N = 7$ for exercise D1 and $N = 6$ for exercise D2) of interdependence strength for the Dynamics dimension, per exercise and experimental condition. Values with an asterisk denote a statistically significant difference between ensemble and solo.

The results are quite similar to the Intonation case, save for generally higher amounts of interdependence and slightly different results for the Linear Correlation method. All four methods now seem to detect higher amounts of interdependence for the ensemble case, for both recorded exercises. The repeated measures ANOVA test once again showed significant separation between ensemble and solo only for the Mutual Information (D1: $F(1, 6) = 21.022, p = 0.004, \eta^2 = 0.55$, D2: $F(1, 5) = 39.082, p = 0.002, \eta^2 = 0.90$) and the Nonlinear Coupling Coefficient (D1: $F(1, 6) = 28.048, p = 0.002, \eta^2 = 0.66$, D2: $F(1, 5) = 28.409, p = 0.003, \eta^2 = 0.82$).

Timbre

The four interdependence methods were applied on the bow-bridge distance feature as well as the spectral crest feature, as they were extracted from the ensemble and solo recordings of the T1 exercise. Table 4.3 contains the averaged interdependence values for each method and feature; a number of 7 interdependence values were obtained from the windowed analysis.

Method	**Interdependence strength, Timbre (T1) bow-bridge distance**				**Interdependence strength, Timbre (T1) spectral crest**			
	solo		**ensemble**		**solo**		**ensemble**	
	μ	σ	μ	σ	μ	σ	μ	σ
Pearson correlation	0.676	0.822	0.642	0.886	0.035	0.182	0.158	0.397
Granger causal density	0.009	0.004	**0.013***	0.005	0.002	0.002	0.008	0.010
Mutual information	1.070	0.247	1.057	0.128	0.170	0.033	**0.341***	0.138
Nonlinear coupling coefficient	0.617	0.085	0.633	0.062	0.154	0.056	**0.231***	0.076

Table 4.3: Mean (μ) and standard deviation (σ, $N = 7$) of interdependence strength for the Timbre dimension, per feature and experimental condition. Values with an asterisk denote a statistically significant difference between ensemble and solo.

For the case of bow-bridge distance, the results are different from the two previous cases. Neither Linear Correlation nor Mutual Information show higher amounts of interdependence for the ensemble condition (although it can be argued that the differences between ensemble and solo are quite small compared to the previous cases). Causal Density and Nonlinear Coupling show higher values for the ensemble condition, although the repeated measures ANOVA test confirms a statistically significant difference only for the Granger Causality measure ($F(1, 6) = 7.628, p = 0.036, \eta^2 = 0.76$).

The results on the spectral crest feature are a return to the pattern observed in previous performance dimensions and especially dynamics. All four methods show higher interdependence values for the ensemble condition, with statistical significance for the Mutual Information ($F(1, 6) = 13.608, p = 0.01, \eta^2 = 0.63$) and Nonlinear Coupling Coefficient methods ($F(1, 6) = 6.747, p = 0.041, \eta^2 = 0.66$).

Timing

The four interdependence methods were applied on the tempo curve feature as it was extracted from the ensemble and solo recordings of exercises R1 and R2. Table 4.4 contains the overall interdependence values for each method and feature;

as the tempo curve feature is not windowed, we report solely one value per method and experimental condition.

Method	Interdependence strength, Timing (R1)		Interdependence strength, Timing (R2)	
	solo	ensemble	solo	ensemble
Linear correlation	0.718	0.983	0.523	0.830
Causal Density	0.362	0.697	0.160	0.462
Mutual Information	0.534	1.207	0.292	0.715
Nonlinear Coupling Coefficient	0.066	0.112	0.026	0.063

Table 4.4: Interdependence strength for the Timing dimension, per exercise and experimental condition.

All four methods show higher amounts of timing interdependence for both exercises, with the amount of interdependence in the ensemble recordings being more than double the amount of solo interdependence (with the exception of Linear Correlation). The lack of sufficient data means that a test for statistical significance for each interdependence method cannot be carried out in this case.

4.2.1.3. Discussion

General remarks

The obtained results of our statistical analysis suggest that, at least for the simple exercises studied here, it is feasible to correctly discriminate between recordings where no interaction between the performers exists, and recordings where some interdependence can be safely assumed to exist, for all of the studied performance dimensions.

The interdependence methods capable of detecting nonlinear interactions appear to be the most suitable for such an analysis, especially for the case of the intonation and dynamics performance dimensions. Based on the obtained results, we offer two potential explanations for this finding.

The first potential explanation lies with the properties of the extracted features. As it can be observed in Figure 4.7, the features for which linear methods performed poorly (pitch deviation, sound intensity, spectral crest) feature abrupt discontinuities at note boundaries. For the pitch deviation and sound intensity features, these discontinuities are artificially introduced during the computation of the features, while for the spectral crest the discontinuities are inherent to the characteristics of the feature, as the audio spectrum is bound to become noisier during note transitions. For the case of Pearson correlation, these discontinuities produce outlier points which make the computation of correlation less reliable. On the other hand, Granger causality uses multivariate autoregressive models in order to perform its regression analysis, and the existence of discontinuities make the features difficult to approximate using autoregressive processes.

Going beyond the properties of the time series features, a second potential explanation is that the dynamics of the relationships between musicians are non-linear themselves. This seems particularly plausible for the case of the intonation dimension, where the choice of tuning temperament -and therefore the choice of playing a note flatter or sharper- depends on the interval between each performed note and the notes in the scores of other members of the ensemble. Positive pitch deviation in one note might be followed by negative pitch deviation in the subsequent note, in a manner that is not easily predictable from the score using computational means. A way to assess whether this explanation is valid would be to carry out recordings of much lower-level musical tasks (such as a simple musical scale). Another option would be to use a different tuning temperament system for each note interval rather than using equal tempered intonation as a reference; this way, absolute pitch deviation could be used as a metric that captures overall deviation from the chosen temperament, which would produce a linear approximation of intonation behavior.

Intonation

For the analyzed intonation exercise, it should be additionally noted that intonation choices (in terms of how intervals are performed) in the solo condition are bound to be quite different from intonation choices in the two ensemble conditions, mainly due to the lack of a common goal (i.e. consonant chords) in the solo condition. Although not part of our experimental design, an additional solo recording that is carried out after the ensemble recording could perhaps shed more light on the effect of such choices on intonation interdependence, as well as investigate the effects of auditory imagery on solo performance.

Although we did not find a statistically significant difference between the two ensemble recordings of the intonation exercise (without annotations and with annotations), the nonlinear methods reported slightly higher interdependence values for the second recording; whether this is a result of the order of the recordings, the existence of the annotations or chance, we cannot say; more repetitions of the two recordings are necessary to answer the question.

Dynamics

Regarding the analyzed dynamics exercise, the results are nearly identical to the case of the intonation exercise in terms of the methods showing significant difference between ensemble and solo, although all four methods report higher interdependence values for the ensemble recording. The latter can be attributed to the nature of the dynamics performance dimension itself, since deliberate fluctuations in dynamics can be expected to occur in a similar way across the members of an ensemble (as opposed to intonation fluctuations which depend on the note interval).

In our analysis, we detrend the sound intensity time series between note boundaries in order to reduce the effect of the score and make within-note fluctuations of sound intensity more prevalent. Ideally, a combination of different extracted features and interdependence methods that does not require such a processing of the signal would be desirable, as it would allow us to study the phenomenon in a more direct way.

Timbre

In the analyzed timbre exercise, the results obtained present a mixed picture. For the bow-bridge distance feature, we have only obtained significant separation between ensemble and solo using the Granger causality measure. The fact that artificially introduced discontinuities do not exist in the bow-bridge distance time series, this result appears reasonable; after all, Granger causality has been previously used with success on similar bowing movement data from string instrument performances (D'Ausilio et al., 2012).

At the same time, more sophisticated bowing features could be used to describe timbre, such as the bow force, velocity, angle, or a normalized variant of the bow-bridge distance feature divided by the finger position of the performer. In this analysis, it was our intention to study timbre in a manner separate from other performance dimensions such as dynamics and intonation; however, perhaps by doing so we use features that describe timbre in a limited way. The same holds for the spectral crest feature; although significant separation between solo and ensemble was observed with the nonlinear methods, there is a host of spectral timbre descriptors in the Music Information Research literature that can be assessed for their capability to capture interdependence while providing a more complete image of the timbre of the produced sound.

Timing

Timing interdependence computed for the two rhythm exercises showed clear differences between solo and ensemble recordings, although the lack of repetitions did not allow us to perform a statistical analysis to investigate whether the observed differences are significant. Most of the existing studies on timing deal with ensemble synchronization on a level of individual notes, based on the thoroughly studied phase correction and period correction phenomena and their mathematical background. A joint analysis of temporal behavior data using both this approach as well as the score-level approach that is followed in our study could highlight their differences or common ground and provide broader description of joint musical action in terms of temporal coordination. Similarly, the inclusion of these methods in the work presented here could help in bridging the conceptual gap between the concepts of synchronization and interdependence.

Closing remarks

In our review of the existing literature (see Chapter 2.1.2.3), we discussed how performance dimensions such as intonation, dynamics, and timbre are relatively underrepresented in the ensemble performance literature. The results we have obtained show that it is feasible to consider including these dimensions in future studies, and their inclusion could expand and consolidate aspects of our approach for the cases where our data were not enough to offer strong conclusions.

Two important aspects of musical interdependence that are not addressed in this study are (i) the fluctuation of interdependence strength along time, and (ii) the interpersonal relationships among the musicians and the roles they imply. In its current state, our methodology allows for the investigation of both time-varying estimations (through the use of a windowed analysis) as well as separate estimations for each pair of performers, even though we have not dealt with such questions so far. We believe that the release of our experimental recordings in the form of a public dataset will encourage carrying out extended research on these two topics.

4.2.2. Relative interdependence across performance dimensions

4.2.2.1. Overview

Up to this point, we have only discussed interdependence results in an ensemble versus solo context - i.e. without drawing any comparisons across the different exercises for the same performance dimension. Such a comparison would indeed be valuable in assessing the capability of the interdependence methods for not only detecting interdependence, but also quantifying how it varies across different recorded exercises as a way to gain insights on the shared goal of the ensemble.

However, the score of each exercise contains different numbers of notes, note durations, and types of relations between the parts of the musicians. For that reason, the interdependence estimates of different ensemble recordings are not directly comparable, as the overall amount of interdependence that is estimated for each piece may vary regardless of the actual amount of cooperation that exists among the performers.

For this reason, we make use of the interdependence estimated from the solo recordings as a baseline to which the ensemble interdependence can be compared. More specifically, we subtract the estimated solo interdependence from the estimated ensemble interdependence in order to obtain a relative measure, the amount of "interdependence gain" that is observed as a result of joint action. Given how there is a single solo and ensemble recording for each exercise, the comparisons reported in this subsection are difficult to generalize or statistically test. Besides the fact that the actions of the musicians are not always deliberate or a consequence of interdependence, deviations across repeated performances of the same exercise should be expected.

In order to draw a fair comparison, we also carry out this analysis over a wide range of values for the window size parameter (for intonation, dynamics and timbre) and the smoothing window width parameter (for timing); this way, we can investigate the effect of these parameters on the estimated interdependence. The Mutual Information method was chosen for the repeated estimation of relative interdependence for a range of window sizes (for the intonation, dynamics and timbre dimensions) and smoothing windows (for the timing dimension), due to its rapid computation time as well as its good performance in terms of quantifying interdependence on the analyzed data.

4.2.2.2. Results

Intonation
For the intonation dimension, we compute the difference in Mutual Information between solo and ensemble recordings, estimated using the pitch deviation feature over a range of analysis window sizes (from 4 to 20 beats). Figure 4.10 shows the obtained results.

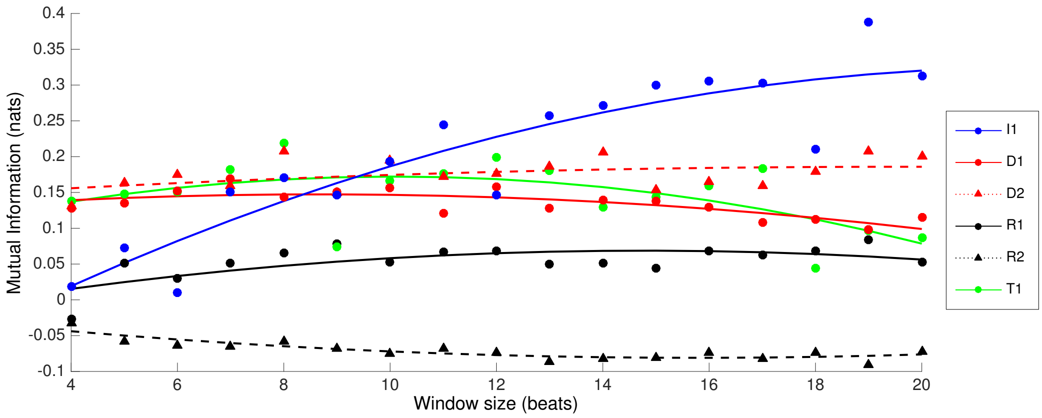


Figure 4.10: Interdependence gain for the intonation dimension, for different exercises and window sizes. I1 - Intonation exercise, D1 - Dynamics exercise nr.1, D2 - Dynamics exercise nr.2, R1 - Rhythm exercise nr.1, R2 - Rhythm exercise nr.2, T1 - Timbre exercise.

As it can be observed, the I1 exercise gradually increases in relative interdependence gain as the window size becomes larger, up to the point where it shows notable interdependence gain in comparison to the rest of the exercises. In other words, the larger the temporal context provided, the clearer the importance of Intonation interdependence becomes for the I1 exercise. Besides I1, the exercises with the next highest interdependence gain appear to be D1, D2 and T1. On the other hand, we can also observe that the curve of the R2 exercise dips below the zero mark, signifying that interdependence for the solo condition was actually higher than the ensemble condition; we provide some comments on this finding and similar occurences for the rest of the performance dimensions in the discussion section below.

Dynamics

For the dynamics dimension, we compute the difference in Mutual Information between solo and ensemble recordings, estimated using the detrended sound intensity feature over the same range of analysis window sizes. Figure 4.11 shows the obtained results.

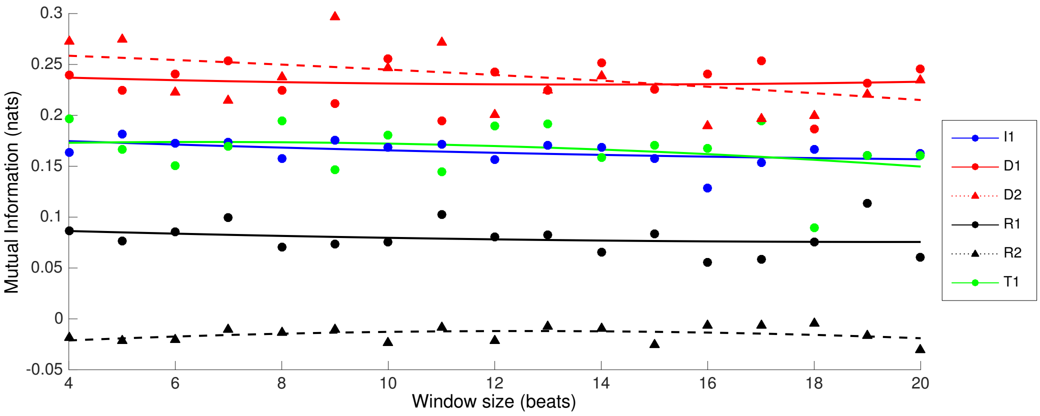


Figure 4.11: Interdependence gain for the dynamics dimension, for different exercises and window sizes. I1 - Intonation exercise, D1 - Dynamics exercise nr.1, D2 - Dynamics exercise nr.2, R1 - Rhythm exercise nr.1, R2 - Rhythm exercise nr.2, T1 - Timbre exercise.

The D1 and D2 exercises consistently demonstrate the highest amounts of interdependence gain, followed closely by the T1 and I1 exercises. Moreover, the window size does not appear to greatly affect the gain of interdependence for either of the exercises. In a similar vein to the Intonation case, the two tempo-based exercises (R1 and R2) demonstrate low and negative interdependence gain, respectively.

Timbre

For the timbre dimension, we compute the difference in Mutual Information between solo and ensemble recordings, estimated using the spectral crest feature. Figure 4.12 shows the obtained results.

The T1 exercise is shown to sustain a consistent amount of interdependence gain across different frame sizes along with the D1 and D2 exercises, which also show comparable amounts of interdependence gain. Exercises R1 and R2 demonstrate low amounts of interdependence gain, while exercise I1 shows negative amounts of interdependence gain.

Timing

For the timing dimension, we compute the difference in Mutual Information between solo and ensemble recordings, estimated using the tempo curve feature. We remind the reader that timing interdependence is not computed in a windowed fashion due to the small amount of data points in each tempo curve feature; we thus repeated the Mutual Information estimation over a different parameter, the

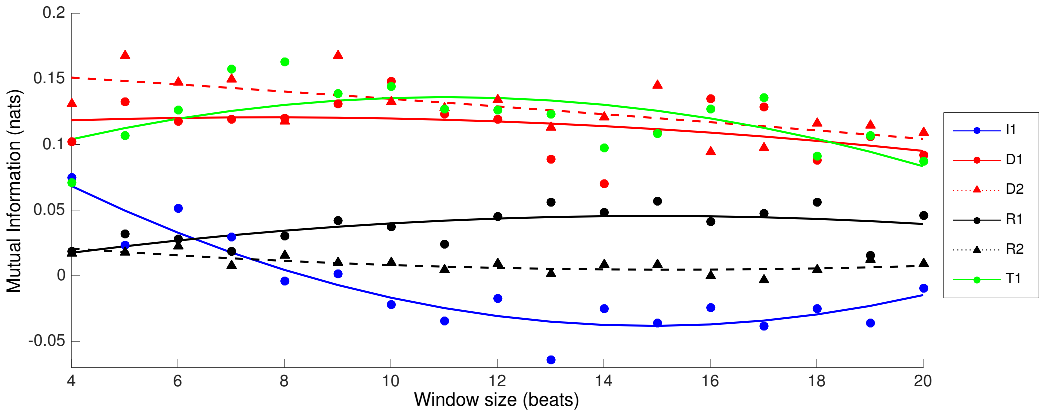


Figure 4.12: Interdependence gain for the timbre dimension, for different exercises and window sizes. I1 - Intonation exercise, D1 - Dynamics exercise nr.1, D2 - Dynamics exercise nr.2, R1 - Rhythm exercise nr.1, R2 - Rhythm exercise nr.2, T1 - Timbre exercise.

width smoothing of the smoothing window (from 1 to 10 beats). Figure 4.13 shows the obtained results.

Exercise R1 consistently demonstrates the highest amount of interdependence gain, while exercise R2 demonstrates this behavior from a smoothing window of 5 and on. Increasing the smoothing window appears to lead to an increase in interdependence gain for both exercises.

In addition to timing interdependence for all exercises, we computed two additional statistics regarding timing: the Mean Absolute Asynchrony between each pair of simultaneous notes in a score, and the Mean Note Duration. For each note i in the score of the ensemble recording with a total number of notes I, we obtain the note onset time $t^i_{0,x}$ and note offset time $t^i_{1,x}$, where $x \in [1, 4]$ represents violin 1, violin 2, viola and cello, respectively. The Mean Absolute Asynchrony $\overline{A}$ is computed as the average asynchrony between all performer pairs for all performed notes as follows:

$$\overline{A} = \frac{1}{I} \sum_{i=1}^{I} \frac{\sum_{\substack{x=1 \\ x<y<4}}^{4} |t^i_{0,x} - t^i_{0,y}|}{6} \tag{4.7}$$

The Mean Note Duration $\overline{D}$ is computed as the average duration of all notes over all performers:

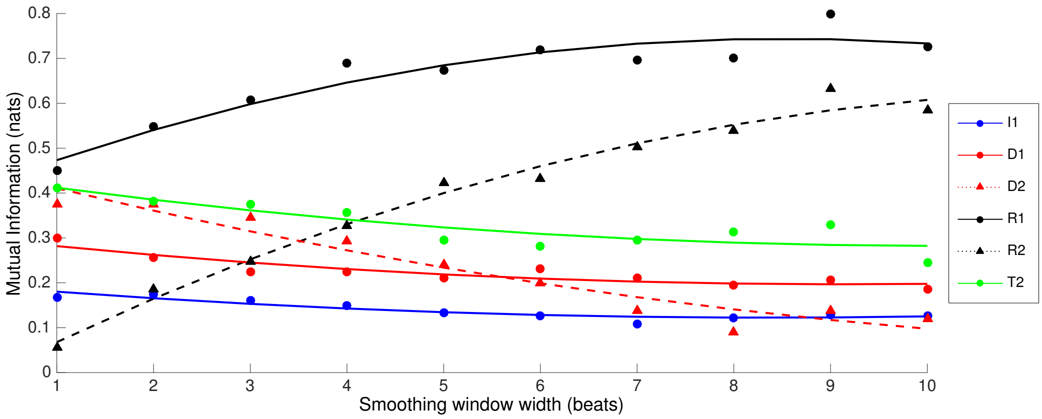


Figure 4.13: Interdependence gain for the timing dimension, for different exercises and smoothing window sizes. I1 - Intonation exercise, D1 - Dynamics exercise nr.1, D2 - Dynamics exercise nr.2, R1 - Rhythm exercise nr.1, R2 - Rhythm exercise nr.2, T1 - Timbre exercise.

Exercise ID	**D1**	**D2**	**I1**	**R1**	**R2**	**T1**
MAA (seconds)	0.100	0.091	0.114	0.042	0.037	0.118
MND (seconds)	4.535	4.419	6.555	0.939	0.572	4.624

Table 4.5: Mean Absolute Asynchrony (MAA) and Mean Note Duration (MND) for each exercise.

$$\bar{D} = \frac{1}{I} \sum_{i=1}^{I} \frac{\sum_{x=1}^{4} (t^i_{1,x} - t^i_{0,x})}{4} \tag{4.8}$$

The obtained values of Mean Absolute Asynchrony and Mean Note Duration for each exercise can be seen in Table 4.5.

It can be seen that across all exercises, the asynchrony between musicians can vary from small values (approximately 40 milliseconds, R1/R2) to large values (approximately 120 milliseconds, T1). A correlation analysis between Mean Note Duration and Mean Average Asynchrony revealed a strong positive correlation ($\rho = 0.95$, $p < 0.05$) between them, suggesting that pieces with longer note durations can make asynchrony more "affordable".

4.2.2.3. Discussion

General remarks

Building on the results of the first interdependence analysis detailed in Section 4.2.1, the results of this analysis suggest that it is feasible to not only detect interdependence in ensemble performance, but to also quantify its relative amount for a given performance dimension by comparing across different exercises.

We have seen that in every performance dimension, the relative highest amount of interdependence gain is observed for the exercises whose shared goal is achieving interdependence in terms of that dimension. The results also suggest a divide between exercises focusing on temporal coordination and all other exercises: Rhythm exercises demonstrate lower or negative amounts of interdependence gain for the intonation, dynamics and timbre dimensions, while the opposite is seen for the timing dimension.

The results hint at how interdependence is a quantity that varies with the goal of the performed exercise. While promising, the findings are not yet conclusive due to the limited amount of data that we have collected. Recording more repetitions for each experimental condition would help to assess whether findings are due to deliberate action rather than a result of the innate variability of music performance. Additionally, we can only draw conclusions based on the relative results between exercises, rather than the absolute interdependence gain values obtained. Therefore, this analysis serves as an exploration and potential direction for further research rather than a pursuit of definitive conclusions.

Intonation

Regarding the comparison across exercises for the intonation dimension, we have noted two interesting findings: the relative increase of interdependence gain for the I1 exercise along larger analysis windows, and the negative interdependence gain for the R2 exercise. Regarding the first finding, a possible explanation is the fact that intonation behavior depends on the intervals between the performed note and the notes of the rest of the ensemble. Thus, in order to obtain a clear picture on intonation interdependence, a window size that covers at least two bars -and therefore two consecutive sets of intervals- is necessary. Regarding the second finding, a potential explanation can be provided by the score of exercise R2 (see Appendix A); the exercise consists of the same two notes for each musician, performed at different degrees of rhythmic syncopation. In the solo condition, the performers have no external perturbations and can therefore maintain a much more steady and predictable intonation behavior, which could be interpreted as coordinated action by the interdependence measures.

Dynamics

A similar result is observed for the dynamics dimension, again for the case of exercise R2. Beyond the characteristics of the exercise and its effect on the solo dimension, another explanation can be proposed in relation to the nature of the rhythm exercises themselves and the shared goal of the ensemble; namely, that the performers playing together are concentrating their efforts on correct timing and exercising less effort to achieve other types of interdependence. This, in combination with more predictable behavior in the solo recording, could be either verified or dismissed with the analysis of additional repetitions for the solo and ensemble recordings of the each exercise.

Timbre

Regarding the results for the timbre dimension, we have observed comparably high values of interdependence gain for the T1, D1 and D2 exercises. Exercise R2 yields close to zero interdependence values for the R2 exercise, and negative interdependence gain values for the I1 exercise save for the first few window sizes. Again, although we suspect the more predictable behavior in the solo condition to be responsible for this finding, it is difficult to draw a definite conclusion. Additionally, as discussed in the previous section as well as Chapter 2, timbre is a quality of the performance that is challenging to quantify, and the current results show that there is ample room for improvement in our approach.

Timing

For the timing dimension, we have observed a clear increase of interdependence gain along with the size of the smoothing window - a result that, as in the case of intonation, suggests that studying the data using a larger temporal context leads to a clearer view at the underlying goal of the exercise. In support of this, the rest of the exercises maintain a steady level of interdependence gain. In addition, the rest of the exercises also showed a larger overall asynchrony between note onsets (as seen in Table 4.5), which results in larger differences between the tempo curve features. The fact that the Dynamics, Intonation and Timbre exercises sustain high amounts of interdependence in most other performance dimensions despite the large asynchronies supports the notion that synchronization and interdependence are two separate qualities, each describing a different aspect of ensemble performance.

Closing remarks

As is usually the case with quantitative research that combines different algorithms and feature extraction techniques, each step requires technical decisions regarding the selection of appropriate parameters, the impact of which on the final result is not always straightforward to predict and often requires iterative computations in order to assess; the impact of analysis window size and smoothing window width are a clear example of this. Another example can be seen in Figure 4.9; although in the previous study we did not discover a notable impact of the Nonlinear Coupling coefficient's parameters on the results, such iterative computations are necessary in order to assess their impact.

Both the analysis presented both here as well as the statistical comparison between solo and ensemble recordings reported in Section 4.2.1 requires the acquisition of reference recordings where (ideally) no interaction between the performers takes place, represented by the solo recordings in our methodology. Data collection under these conditions is a complicated and costly process that can only be carried out under experimental conditions. On the contrary, capturing the performance of a quartet under natural conditions is relatively straightforward and commonplace even within the recording industry. It is therefore desirable to eliminate the need for such reference recordings in order to make our methodology more easily applicable. An interesting direction that can potentially achieve this is using surrogate time series (Schreiber and Schmitz, 2000) to generate our reference time series directly using features from the ensemble recordings. Current techniques involve either transforming the original data to generate surrogate data that lack interdependent behavior, or fitting a model to the data and then changing the parameters of the model in order to eliminate any interdependent components. We carried out initial steps in this direction without yet obtaining satisfactory results; the pursuit of this direction in the future could greatly simplify the application of our methodology on data captured outside controlled conditions.

4.3. Aural perception of interdependence

As we saw in Section 4.2.1, using computational means to detect evidence of interdependence between the performers has proven to be a challenging yet feasible task, at least for the simple exercises we analyzed. Our motivation behind this study is to assess whether the same can be achieved by human listeners for more complex musical compositions; such an investigation can help us understand which aspects of interdependence are most salient from the point of view of the listeners, as well as identify the factors that affect their perception.

Previous work on this subject is limited. Glowinski et al. (2012) carried out an experiment where subjects were asked to decide whether a recorded segment was performed solo or as an ensemble by observing only the first violinist of a string quartet ensemble. Besides the perceived performance condition, the subjects also rated the expressivity of the musician and the expressed emotions of the performance, while describing which of the body features of the musician (head motion, arm motion, etc.) they focused on in order to make their assessment. Results did not show significantly different assessments for the solo and ensemble conditions, although the expressivity and expressed emotion ratings showed some significant interaction with the two conditions (solo, ensemble); the lack of significant differences between solo and ensemble ratings can be partially attributed to the fact that subjects only observed the behavior of the first violinist rather than the entire ensemble.

Listening experiments have been employed in similar tasks where a binary decision on performed music is required. Examples include judging whether a recorded performance was composed or improvised (Lehmann and Kopiez, 2010), whether different excerpts had been played by the same performer (Gingras et al., 2011), or whether the audio and video channels of recorded performances are synchronized (Bishop and Goebl, 2014). Finally, listening experiments have also been used to evaluate the simulation of an orchestral violin section from a single recording (Gingras et al., 2011).

Our aim in this study is to assess how reliably can human listeners detect evidence of musical interdependence when listening to recorded performances of an ensemble. This is achieved by carrying out a listening experiment where listeners compare real string quartet recordings to artificially synchronized solo recordings of the same piece. We utilize a subset of the short piece excerpts recorded in the main experiment, and investigate how the judgments of the listeners are affected by the type of piece as well as the listeners' own background.

4.3.1. Overview

Materials
The recordings used for the listening experiment consist of five piece excerpts; Table 4.6 contains a summary of their ID used within the experiment, the piece excerpt from the QUARTET dataset that they correspond to, and their duration:

Artificial synchronization of solo recordings
Given that the recordings were carried out without a metronome, it was necessary to artificially synchronize the solo recordings in a similar way to what was done in Section 4.1.1.1. Moreover, since our goal was to assess whether listeners can detect musical coordination based on factors other than rhythmic synchronization,

ID	Corresponding piece excerpt	Duration
P1	Phrasing piece excerpt 1 (PP1)	00:58
P2	Dynamics/Intonation piece excerpt 2 (DP2/IP2)	00:46
P3	Dynamics/Unity of Execution piece excerpt 3 (DP3/UP3)	00:36
P4	Rhythm piece excerpt 1 (RP1)	00:42
P5	Rhythm piece excerpt 3 / Phrasing piece excerpt 2 (RP3/PP2)	01:21

Table 4.6: Summary of the excerpts used for the listening experiment.

it was also necessary to ensure that the solo recordings had exactly the same note onset/offset times as the ensemble recordings.

We utilized a transient-preserving time scaling algorithm (Bonada, 2000) to apply a note-by-note time stretch to the solo recordings using the ensemble recordings as reference: for each individual instrument, the audio signal is partitioned using the note onset times as anchor points; then, the duration of each solo note is altered to match the duration of the corresponding ensemble note in the score; finally, the solo waveform is shifted to coincide with the ensemble waveform.

We carried out a pilot test to assess whether any audible artifacts are introduced by this procedure using music technology researchers as subjects, without encountering any. Earlier variants of this time-scaling algorithm have been also used in listening experiments without introducing any significant bias (Honing, 2006).

Post-processing

Given that bridge pickup recordings have a certain "nasal" quality, all four pickup (bridge vibration) signals were respectively convolved with body impulse responses obtained from the same instruments used in the experiment recordings. Measurements of force and velocity using a miniature impact hammer and a vibrometer were obtained to create models of driving-point bridge admittances, from which the body impulse responses can be derived (Maestre et al., 2013).

In order to reconstruct the stereophonic image of a string quartet, the four recordings in each excerpt were panned from left to right as follows: violin 1 (60% left), violin 2 (20% left), viola (20% right), cello (60% right). Finally, the gains applied to the audio signal of each instrument were manually set using stereo recordings of each excerpt as reference; the same gain for each instrument was applied to all recordings.

Experiment
The listening experiment was carried out through an online survey system. Each subject was asked to use headphones in order to ensure similar listening conditions. Before listening to any recordings, the following personal information was gathered:

1. Age
2. Gender
3. Amount of (formal or informal) musical training (*None, Up to 2 years, between 3 and 5 years, more than 5 years*)
4. (Conditional to Training) Experience with bowed string instruments

After this step came Phase 1 of the experiment: the subject listened to the five recording pairs (solo and ensemble) in random order within the experiment (solo first or ensemble first), but the same order across all subjects. It is important to note that, at this time in the experiment, the subject was not aware that only one of the recordings is from a "real" ensemble. The subject was tasked with listening to each pair of recordings and comparing them in terms of *Quality of performance* and *Degree of coordination*; there was also the option of considering both recordings equal.

In Phase 2, the subject was then informed that one of the recordings is performed by an ensemble while the other originates from artificially synchronized solo performances. Then, the subject listened to the same five recording pairs again, this time with the task of choosing the recording he/she believed to be the real ensemble recording. Similarly to Phase 1, the subject could answer "I am unable to decide". Finally, a comments form was provided for each excerpt where the subjects could specify what helped them make their decision.

4.3.2. Results

We analyzed the responses of 74 volunteer subjects (51 males). The mean age of the subjects was 32 years old ($\sigma = 11$). 39 subjects had received more than 5 years of musical training, while 8 subjects had experience with bowed string instruments.

Figure 4.14 shows which recording was rated with a higher "performance quality" per excerpt across all subjects, while Figure 4.15 shows which recording was rated with a higher "degree of coordination".

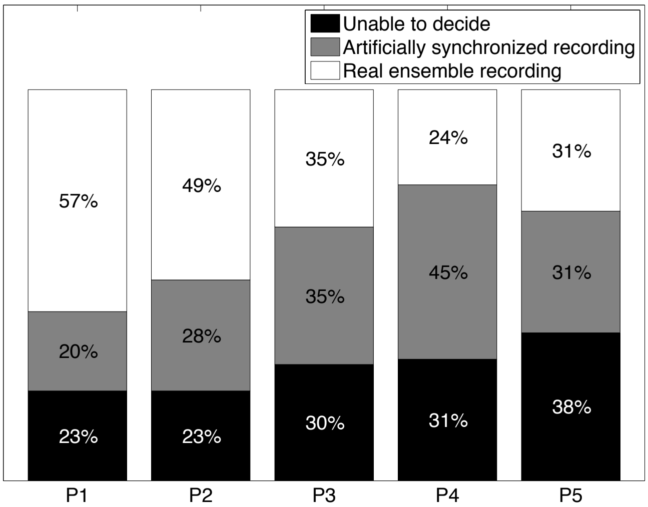


Figure 4.14: Collected responses for all subjects on the recording featuring higher performance quality (Phase 1). See Table 4.6 for the meaning of the piece excerpt IDs (Px).

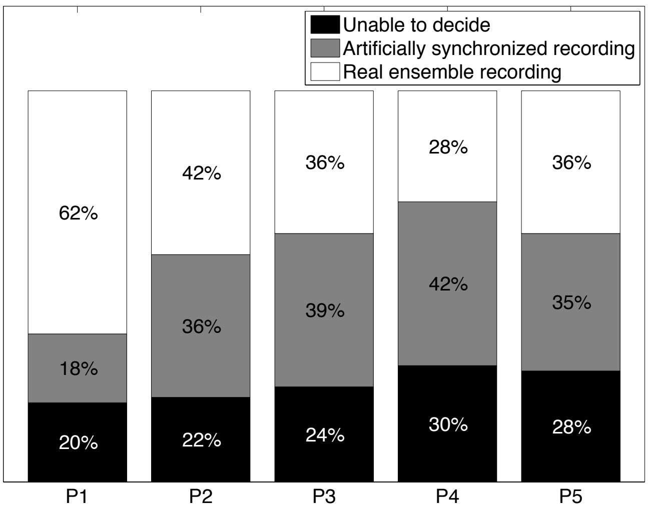


Figure 4.15: Collected responses for all subjects on the recording featuring higher performance coordination (Phase 1). See Table 4.6 for the meaning of the piece excerpt IDs (Px).

It can already be seen from the above two figures that each excerpt elicits a different response from the subjects. Especially the last two excerpts seem to be the most difficult to compare; given that we selected those two excerpts as examples of rhythmic coordination, it seems plausible that by making the solo and ensemble recordings identical in terms of note onsets and offsets we are eliminating differences in the aspect of the performance on which the musicians were most focusing on.

Another observation that can be made from the above figures is that the ratings of the subjects for "performance quality" and "degree of coordination" appear to be in relative agreement; this was confirmed by measuring the Spearman rank correlation coefficient between these two factors per excerpt; the obtained ρ values are as follows: P1: 0.84, P2: 0.73, P3: 0.72, P4: 0.66, P5: 0.71 ($p < 0.001$ for all cases).

Regarding Phase 2 of the experiment, Figure 4.16 shows which recording was chosen as the real quartet recording across all subjects, per excerpt.

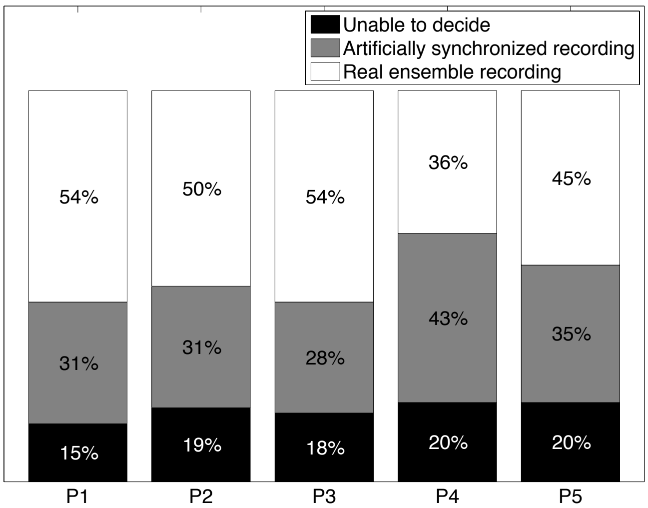


Figure 4.16: Collected responses for all subjects regarding the recording featuring "real" ensemble performance (Phase 2). See Table 4.6 for the meaning of the piece excerpt IDs (Px).

Again, it can be seen that listeners encounter difficulties in detecting the real quartet recording, with some piece excerpts showing higher accuracy than others in the same way as in Phase 1.

So far, we have not investigated the effect of musical training on the responses of the subjects; moreover, although we have seen that different excerpts provide varying results, the effect of each excerpt remains to be seen. In order to investigate these factors, we performed a logistic regression on the binary outcome of each comparison (1 for the cases where the real ensemble recording was chosen, and 0 otherwise). The results for each comparison (performance quality, performance coordination, final assessment) can be seen in Tables 4.7, 4.8 and 4.9, respectively.

Coefficient	**Estimate**	**Std. error**	**p-value**
Excerpt P1	-0.404	0.320	0.207
Excerpt P2	-0.754	0.324	**0.020 ***
Excerpt P3	-1.361	0.341	<**0.001 ***
Excerpt P4	-1.932	0.370	<**0.001 ***
Excerpt P5	-1.562	0.350	<**0.001 ***
Training	0.277	0.108	<**0.010 ***
String exp.	1.409	0.399	<**0.001 ***

Table 4.7: Logistic regression results for performance quality.

Coefficient	**Estimate**	**Std. error**	**p-value**
Excerpt P1	-0.285	0.319	0.370
Excerpt P2	-1.161	0.330	<**0.001 ***
Excerpt P3	-1.402	0.337	<**0.001 ***
Excerpt P4	-1.794	0.354	<**0.001 ***
Excerpt P5	-1.402	0.337	<**0.001 ***
Training	0.363	0.107	<**0.010 ***
String exp.	0.667	0.370	0.071

Table 4.8: Logistic regression results for performance coordination.

Coefficient	**Estimate**	**Std. error**	**p-value**
Excerpt P1	-0.568	0.309	0.066
Excerpt P2	-0.738	0.311	<**0.018 ***
Excerpt P3	-0.568	0.309	0.067
Excerpt P4	-1.314	0.325	<**0.001 ***
Excerpt P5	-0.964	0.315	**0.002 ***
Training	0.356	0.101	<**0.010 ***
String exp.	0.068	0.357	0.847

Table 4.9: Logistic regression results for the final assessment.

From the above results one can observe that the amount of musical training has a significant positive effect on the outcome; that is, subjects with higher amounts of musical training tend to be more accurate. Regarding experience with bowed string instruments, we could detect a significant positive effect only on the assessed performance quality; the small amount of subjects with string experience (8 out of 74) makes conclusive results difficult to achieve, and we believe that a more thorough investigation of the matter is called for.

Regarding the excerpt type, we can observe that excerpts P2, P4 and P5 seem to have the most significant effect on the ratings of the subjects, at least for the final decision in Phase 2; for Phase 1 decisions, excerpts P2 to P5 all seem to significantly affect the ratings of the subjects. We could not find any significant interaction between the regression coefficients; the skewed distribution of some variables (such as experience with string instruments) makes the assessment of interaction between coefficients difficult.

4.3.3. Discussion

This study investigated the capability of listeners (or lack thereof) to distinguish between real and artificially synchronized recordings of string quartet excerpts, once the aspect of rhythmic synchronization has been eliminated. The results suggest that the task is challenging yet achievable for human listeners.

The aural perception of interdependence proved easiest for piece excerpts with a focus on phrasing, dynamics and intonation, while piece excerpts with a focus on rhythm proved more difficult to distinguish. This suggests that synchronization, while of high importance, is not the only aspect of ensemble performance that is reflected through the acoustic result.

Ratings obtained by the subjects showed a positive correlation between the perceived performance quality and degree of coordination, showing that a perception of interdependence among the performers tends to be associated with higher quality in the performance and vice versa.

It has also been seen that musical training can improve the capability of the listeners to perceive interdependent behavior and correctly discriminate between real and artificially synchronized performances, while the characteristics of the piece (in terms of the performance dimensions it focuses on) was also seen to be an important factor in perceiving interdependence.

The obtained results are promising and call for additional studies on the topic. A more diverse selection of musical pieces (including the rest of the piece excerpts recorded as part of our main experiment) as well as more participants with string performance experience could be included in further refinements of the experiment, while a detailed analysis on computational methods of interdependence, score-level features, and their relation to the judgment of the listeners can be investigated.

Chapter 5

CONCLUSION

In a musical ensemble, the performers interact and influence each other's actions in several aspects of the performance simultaneously in order to achieve a shared goal. In this dissertation we have focused on this phenomenon, the *interdependence* between the performers, with the goal of designing a computational data-driven methodology to observe and study interdependence, using a string quartet ensemble as a case study. Combining elements from multiple disciplines, this methodology involves the acquisition of multimodal performance data under experimental conditions, the extraction of numerical features that describe the performance in terms of distinct musical dimensions (intonation, dynamics, timbre, timing), and the computational estimation of interdependence for each performance dimension individually.

Through our review of the relevant scientific background on music ensemble performance we have noted a general lack of research on ensembles in comparison to the volume of research on music performance in general; and while a surge in ensemble studies has occurred in the last 10 to 15 years, the majority focuses on the synchronization between performers without dedicating much effort on other performance dimensions.

Motivated by this fact, we conducted a large-scale experiment where the performance of a string quartet was captured using multiple data modalities under controlled conditions. The musicians performed a broad array of exercises whose central connecting characteristic was their focus on achieving interdependence in terms of different dimensions of the performance, as well as excerpts of musical pieces that serve as more natural and less controlled examples of musical interdependence. We implemented a methodology for the multimodal acquisition and processing of audio, motion capture, and video data. All of the recorded data have been released into the public domain as a freely available dataset along with some of the tools used to create it, with the intention of advancing research on the interdependence of musical ensembles.

We have analyzed a subset of the recordings, combining audio and motion feature extraction techniques with time series analysis methods to investigate whether interdependence can be measured using computational means in terms of four performance dimensions: intonation, dynamics, timbre, and timing. By comparing recordings of the ensemble with artificially synchronized solo recordings of each individual musician, we have been able to demonstrate that interdependence is indeed measurable for each of the performance dimensions, at least for simple exercises such as the ones we analyzed. Moreover, it was seen that it is also feasible to compare the amount of interdependence across different recordings, and relate the results to the shared goal of the ensemble. An additional perceptual study investigated the capacity of listeners to differentiate between ensemble performances and artificially synchronized solo performances as a function of the listeners' own background and the performance dimension that each recording focused on.

In the following pages we will provide a summary of our contributions, as well as a discussion about the limitations of our approach and the future directions that can be followed from here on. Finally, we will provide some brief closing remarks as a way to conclude this dissertation.

5.1. Summary of contributions

As a result of multidisciplinary research, the contributions of this work are distributed across different topics and are both practical as well as theoretical in nature. In this section we will present them one by one, and discuss their limitations and potential for future work.

5.1.1. Experimental design

In Chapter 3 we proposed an experimental procedure for studying interdependence in a string quartet based on recordings of real performances. The first iteration of our experimental design was evaluated in a pilot study with a violin duet (Papiotis et al., 2011), and subsequently improved and expanded in a large-scale experiment. The final experimental design is centered around two conditions: (i) ensemble, representing normal string quartet performance following rehearsal, and (ii) solo, representing individual performance unaffected by factors that normally influence ensemble performance such as the presence of other musicians or familiarity with the scores of the entire quartet.

The recorded materials are organized in different categories, each focusing on a different aspect of the performance: intonation, dynamics, timbre, rhythm, unity of execution, and phrasing. For each category, we recorded two types of of materials: exercises with a clear shared goal to be achieved by the ensemble, and excerpts of musical pieces where the shared goal is influenced by the category as well as subject to the interpretation of the ensemble. A total of nine exercises and ten piece excerpts were recorded during our main experiment.

Limitations and future work

In our experiment, recordings carried out in the solo condition were designed to be a "ground truth" that was as far removed from the ensemble condition as possible, which proved useful in designing and evaluating our computational methodology. However, it would be also useful to record more variations of the solo condition such as performing alone after the ensemble recording, in order to investigate the effects of auditory imagery on solo performance. Recordings carried out in the ensemble condition are also of a basic nature, representing normal ensemble performance following a short rehearsal. It would be useful to capture additional variations in the ensemble condition, by controlling the type of feedback (auditory/visual) available to the musicians during the ensemble, or even carry out recordings where the musicians are performing a piece together for the first time.

All of the recordings carried out in our main experiment were done without a metronome, which allowed us to observed the behavior of the performers without any external influences. This however made the artificial synchronization of solo features (for the interdependence analysis) and audio recordings (for the listening experiment) necessary in order to compare ensemble and solo recordings. Future repetitions of the experiment can incorporate the use of a metronome in the recordings, and investigate its effect on the estimated interdependence for both ensemble as well as solo.

The exercises recorded in our experiment represent simplified tasks, whose difficulty lies in achieving interdependence rather than performing one's individual task. As such, the stand on a middle ground between controlled experimental conditions and ecological, real-life conditions. Even simpler tasks, such as scales, or repetitive rhythmic tasks (such as the ones present in rhythmic tapping experiments) would provide a closer look to the fundamental processes behind coordination and interdependence in different dimensions of the performance.

Each of the piece excerpts recorded during our experiment consist of short segments, each one associated with specific performance dimensions; as such, they provide a relatively small amount of data when placing focus on ensemble expressivity. Finally, all of our recordings were carried out with a single string quartet of professional musicians and are therefore limited in their potential for generalized assumptions on ensemble performance. Repetitions of the same experimental setup with different quartets could consolidate our findings, or reveal behavior not captured in our experiments.

Longer excerpts, or entire pieces accompanied by an in-depth musicological analysis regarding the association between segments and dimensions of the performance could be carried out, in order to study interdependence in conjunction with expressivity. Besides the types of excerpts, a limitation of our recorded piece excerpts is the fact that the had already been performed by the quartet, and thus recordings carried out in the solo condition were already influenced by the existence of a shared goal between the members of the quartet, as well as auditory imagery. Additional experiments could be carried out, with pieces that do not belong to the quartet's repertoire and are unknown to the performers. This way, the entire learning process ranging from solo performance, through rehearsals, ensemble performance and finally solo performance post-rehearsal would shed light into the gradual development of interdependence over a prolonged period.

5.1.2. Multimodal data acquisition

The experimental procedure detailed in Chapter 3 additionally proposes a technical setup for the simultaneous acquisition of multimodal (sound, movement, video) data in string quartet performance. In this setup, we acquire individual audio from each performer using pickup microphones fitted on the bridge of the instrument, ambient audio using a cardioid microphone and a binaural recording device, instrumental gesture data using both a wired as well as a wireless motion capture system, upper body movement using a wireless motion capture system, and video from the entire quartet. We also implemented a solution to synchronize data recorded by different devices, combining an external sampling clock to prevent sampling drift with the use of a linear timecode signal to synchronize the acquired data streams following acquisition.

In addition to the acquisition of multimodal performance data, we implemented a software solution to carry out score-performance alignment on our recorded data, based on an existing algorithm by Dixon (2005). The code for our score-performance alignment solution is freely available online[1], published under a permissive license as a tool for other researchers working on real performance data.

Limitations and future work

Limitations in the commercial software of the wireless motion capture system, arising from the high amount of reflective markers, led to sporadic crashes during recordings which resulted in missing wireless mocap data for a small subset of the recordings. Additionally, limitations in the storage capacity and data transferring capabilities of the video camera resulted in limited video coverage for several of the recordings.

Wireless marker occlusion as well as artifacts in the detected position of the wireless markers due to IR reflections is a typical problem in camera-based motion capture system, and the effort necessary to clean and process the wireless motion capture recordings proved to be very time consuming. We decided to avoid filtering methods such as smoothing or interpolating marker trajectories for missing frames, and thus provided the cleaned wireless motion caption data with frames of undefined values; there are several software solutions (e.g. the *MotionBuilder*[2] software) that can be used to filter the motion capture data in post-processing, depending on the purpose behind the use of the data.

Although two motion capture systems were used to acquire instrumental gesture data, in this work we have not compared them or evaluated their relative strengths and weaknesses. It is reasonable to assume that the wired system is more intrusive, although the musicians did not appear to be restricted by it. However, a comparative analysis based on the acquired data can be carried out to investigate the degree to which the wireless system eliminates the need for the wired system.

5.1.3. Dataset of string quartet recordings

The experimental multimodal recordings detailed in Chapter 3 represent a valuable resource that can be utilized as material for research in multiple disciplines: ensemble interdependence, ensemble expressivity, instrumental gesture data analysis, music information retrieval, automatic music transcription, audio source separation, multimodal performance analysis, et cetera. From the initial planning of our main experiment, it was our explicit goal to create a rich multimodal dataset on string quartet performance and share it with the rest of the academic community in order to encourage research on ensemble performance; to this end, we recorded data

[1] https://github.com/slowmountain/scorealigner

[2] http://www.autodesk.com/education/free-software/motionbuilder

whose breadth and variety goes beyond the needs of the computational analyses presented in this dissertation.

All of the recordings from our main experiment have been made available online[3] as an openly available multimodal dataset, using a state-of-the-art online repository platform[4] (Mayor et al., 2013). Each recording is in the form of a datapack that contains an .xml file with the structure of the files contained within, and accompanied by textual metadata information.

Limitations and future work

Even though the dataset has been made available, we have not yet taken initiative to draw the attention of the academic community to it, or proposed concrete ways in which the data can be utilized. We are currently in preparation of a scientific article, the purpose of which is to achieve the above challenges.

Additionally, the inclusion of our dataset in Music Information Research challenges like the *Music Information Retrieval Evaluation eXchange* (MIREX[5]) for tasks such as score-performance alignment, multiple fundamental frequency estimation, audio tempo estimation, audio onset detection et cetera would aid both in broadening the reach of music ensemble performance research and research on bowed string instruments across other disciplines.

5.1.4. Computational analysis of interdependence

In Chapter 4, we proposed a computational methodology to measure interdependence in string quartet performance in terms of distinct aspects of the performance, which we termed *performance dimensions* (intonation, dynamics, timbre, timing). Using a subset of the exercises recorded as part of our main experiment, we extracted features in the form of time series that describe performance in terms of each of the four dimensions using both audio as well as instrumental gesture data: pitch deviation as an intonation feature, detrended sound intensity as a dynamics feature, bow-bridge distance and spectral crest as timbre features, and tempo curve as a timing feature.

We applied four interdependence estimation methods based on multivariate time series analysis (pearson correlation, mutual information, granger causality, nonlinear coupling coefficient) to the extracted features in order to estimate interdependence

[3] http://mtg.upf.edu/download/datasets/quartet-dataset
[4] http://repovizz.upf.edu
[5] http://www.music-ir.org/mirex

between the performers in terms of the four performance dimensions. For each exercise, we estimated interdependence from recordings carried out in the ensemble condition, as well as artificially synchronized recordings carried out in the solo condition; the two different estimations of interdependence were statistically compared to assess whether the difference in obtained interdependence values is significant.

The obtained results showed that it is feasible to measure interdependence in terms of each performance dimension, at least for the simple exercises that were analyzed. Moreover, by using the solo recordings as a reference to which the ensemble recordings are contrasted, we carried out a comparison across all exercises per dimension, and found evidence that our methodology is not only capable of detecting interdependence but also quantifying it, showing how the shared goal of each exercise affected the amount of estimated interdependence in each dimension.

To our knowledge, the published studies based on this methodology (Papiotis et al., 2014, 2013b, 2012a,b, 2011) are the first of their kind to simultaneously target intonation, dynamics, timbre and timing in the context of music ensemble performance; in addition, there have been very few studies focusing on intonation and timbre individually. We believe that our methodology provides a novel and promising approach as a basis for future studies on music ensemble performance research, going beyond temporal synchronization and towards joint musical expression.

Limitations and future work

In our computational analysis, we focused on exercise categories that were directly connected to a performance dimension: intonation, dynamics, timbre, and rhythm. We have not yet carried out an in-depth analysis of the remaining two exercise categories - phrasing and unity of execution; doing so could potentially reveal interactions between performance dimensions that we have not yet seen in the results of our computational analyses.

Our objective thus far has been to investigate whether the measurement of interdependence in distinct dimensions of the performance was feasible, as stated in Chapter 1.5.1. Building on the promising results, future research can address more challenging goals, such as studying the temporal variation of interdependence along time, or assessing the interpersonal relationships between the performers in terms of musical roles and leadership. Our proposed methodology provides data for both of these directions, through the windowed estimation of interdependence as well as the use of directional interdependence estimation methods.

Each one of the performance dimensions considered in this work have been studied individually, without considering possible interactions between different dimensions. A meta-analysis of interactions between estimated interdependence values for different performance dimensions can function as a first step in this direction, while a second step could be the recording and analysis of more complex materials that are more meaningful musically.

The materials and experimental conditions considered in this study cover but a small area of the complex phenomenon that is string quartet performance. Our analyses were based on a limited amount of short exercises; longer and more complicated scores could further test the usefulness of our methodology, or reveal interactions between the performance dimensions that were not seen in our studies. Recording more quartets with a different potential for interdependence could validate or challenge the patterns that we have observed so far, while capturing more repetitions of the same performance along a larger time window could reveal information about how interdependence is established during an ensemble's training process.

We represent each performance dimension with time series features extracted from the recorded data. The features that were selected in our approach are but a subsample of a very large set of performance descriptors that can be found in the Music Information Research literature (Serra et al., 2013); and the proposed methodology could be greatly consolidated by including more features and assessing their capacity to capture musical interdependence. This is especially topical for the timbre dimension, as it is particularly difficult to describe the timbre of a musical performance solely with the two features present in our analysis. Beyond the descriptive capabilities of the features themselves, the inclusion of more features with different characteristics regarding stationarity, linearity et cetera could help shed more light on the capability of different interdependence methods in estimating musical interdependence.

The four interdependence methods utilized in our methodology are widely used in several disciplines ranging from Neuroscience to Econometrics and Information theory, and provide some variety regarding directionality and linearity. However, there are several more methods to be evaluated in the future, including variants of the methods utilized in our methodology with different properties. A starting point would be the inclusion of the *Transfer Entropy* and *Coherence* methods discussed in Chapter 2.3.

Our methodology is based around the concept of comparing ensemble recordings with solo recordings that are unaffected from external influences. While this may not appear as an important restriction for the purposes of research, the pre-requisites of the solo recordings utilized in our approach (no previous knowledge of the entire ensemble score, no previous performance with the rest of the ensemble) make them difficult to obtain in more realistic scenarios such as an educational environment or a professional music performance. This limits the applicability of our methodology in real-world tasks. We believe that further research using surrogate time series analysis techniques can help in detecting the components present in our features that contain information on interdependence with greater accuracy, and eliminate the strict requirements of the reference solo recordings.

Thus far we have focused on performance data that either relate to the produced sound of the ensemble, or the bowing gesture movements carried out in order to produce sound. However, our dataset also contains upper body movement data as well as video data, which can be used to study interdependence from a different point of view.

This work has been focused on an ensemble of bowed string instruments. However, it can be reasonably assumed that the presented methodology can be applied to different kinds of ensembles; while ensembles such as wind sections or even singing voice ensembles are an obvious choice, the methodology pertaining to dynamics and tempo can be easily applied to most musical instruments. In fact, the core of our methodology - obtaining numerical representations of the behavior of interacting agents and assessing the interdependence between them - could be useful in studying interactions in social contexts beyond music performance; an example would be the academic discipline of Social Signal Processing (Pantic et al., 2011), where multimodal data originating from human non-verbal behavior are automatically analyzed to study social interaction and derive context-aware causal relationships between interacting agents.

5.1.5. Listening experiment on the aural perception of interdependence

In addition to our computational analysis, we investigated the capability of human listeners to perceive interdependence, using a subset of the piece excerpt recordings. We carried out a listening experiment where the participants were faced with a task similar to the one faced by our computational methodology, i.e. comparing recordings of the ensemble with artificially synchronized solo recordings for the same piece excerpt (Papiotis et al., 2013a).

The obtained results showed that this is an feasible yet challenging task for listeners, whose ability to distinguish between real and artificially synchronized performances was affected both by the piece they were listening to, as well as their own background in terms of musical training. The fact that musically untrained listeners found the discrimination between ensemble and solo recordings challenging, together with the fact that ratings on the quality of performance were shown to be correlated with ratings on performance coordination, raise interesting questions for future studies regarding the motivation behind achieving interdependence in ensemble performance.

Limitations and future work

The limited number of participants in our study, as well as the unbalanced representation of participants with background music and bowed string instrument performance made the formulation of answers to the above questions challenging. Repetitions on the experiment featuring more participants with a balanced representation in background knowledge and experience with bowed string ensembles can provide more conclusive results.

The listening tests were carried out using recordings of the piece excerpts, as the material that is closest to the type of string quartet performance listeners are normally exposed to. This however means that a direct comparison with the results of our computational analysis (which was applied on the exercises instead) is not possible at this point. A combination between our computational methodology and the response of listeners could shed light on the overlap between what is captured by our methodology and what is perceived by listeners.

5.2. Closing remarks

This dissertation has presented a data-driven approach to studying interdependence between the members of a string quartet, focusing on aspects of the performance that - to our knowledge - had not thus far received much attention in the context of joint action, especially for excitation-continuous musical instruments such as bowed string instruments. We regard this approach as exploratory work, a preliminary step towards a deeper understanding of the underlying mechanisms behind ensemble music performance.

It is our belief that, through the outcomes of our research both in terms of experimental data and computational methodology, we have planted a seed for future research to build upon and expand our knowledge on ensemble music performance, a phenomenon whose internal workings are still only partly understood despite its enduring presence.

Bibliography

Abdi, H. (2007). Multiple correlation coefficient. In Salkind, N., editor, *Encyclopedia of Measurement and Statistics*, pages 648–651. SAGE Publications, Thousand Oaks (CA).

Askenfelt, A. (1986). Measurement of bow motion and bow force in violin playing. *The Journal of the Acoustical Society of America*, 80(4):1007–1015.

Askenfelt, A. (1989). Measurement of the bowing parameters in violin playing. II: Bow–bridge distance, dynamic range, and limits of bow force. *The Journal of the Acoustical Society of America*, 86(2):503–516.

Aucouturier, J.-J., Defreville, B., and Pachet, F. (2007). The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *The Journal of the Acoustical Society of America*, 122(2):881–891.

Bach, C. P. E. (1753). *Versuch über die wahre Art, das Clavier zu spielen*. C. F. Kahnt Nachfolger, Leipzig.

Bartsch, M. and Wakefield, G. (2001). To catch a chorus: using chroma-based representations for audio thumbnailing. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 15–18.

Bengtsson, I. and Gabrielsson, A. (1977). *Rhythm research in Uppsala*. Royal Swedish Academy of Music.

Bishop, L. and Goebl, W. (2014). Context-specific effects of musical expertise on audiovisual integration. *Frontiers in psychology*, 5(October):1123.

Bishop, L. and Goebl, W. (2015). When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae*, 19(1):84–110.

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., and Serra, X. (2013). ESSENTIA: an Audio Analysis Library for Music Information Retrieval. *International Society for Music Information Retrieval Conference (ISMIR 2013)*, pages 493–498.

Bonada, J. (2000). Automatic Technique in Frequency Domain for Near-Lossless Time-Scale Modification of Audio. In *Proceedings of the International Computer Music Conference*, pages 396–399.

Cano, P., Loscos, A., and Bonada, J. (1999). Score-Performance Matching using HMMs. In *Proceedings of the International Computer Music Conference ICMC*, volume 1, pages 441–444.

Carabias-Orti, J. J., Rodriguez-Serrano, F. J., Vera-Candeas, P., Cañadas-Quesada, F. J., and Ruiz-Reyes, N. (2013). Constrained non-negative sparse coding using learnt instrument templates for realtime music transcription. *Engineering Applications of Artificial Intelligence*, 26(7):1671–1680.

Chicharro, D. and Andrzejak, R. (2009). Reliable detection of directional couplings using rank statistics. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, 80(2):026217.

Clarke, E. F. and Windsor, W. L. (2000). Real and simulated expression: A listening study. *Music Perception*, 17(3):277–313.

Dannenberg, R. B. and Raphael, C. (2006). Music score alignment and computer accompaniment. *Communications of the ACM*, 49(8):38–43.

D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., Aloimonos, Y., and Fadiga, L. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PloS one*, 7(5):e35757.

Davidson, J. W. (2001). The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. *Musicae Scientiae*, 5(2):235–256.

Davidson, J. W. and Good, J. M. M. (2002). Social and musical co-ordination between members of a string quartet: An exploratory study. *Psychology of Music*, 30(2):186–201.

De Cheveigné, A. and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4):1917–1930.

de Cruys, T. (2011). Two multivariate generalizations of pointwise mutual information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, pages 16–20. Association for Computational Linguistics.

de Mantaras, R. L. and Arcos, J. L. (2002). AI and music: from composition to expressive performance. *AI Magazine*, 23(3):43.

De Poli, G. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research*, 33(3):189–202.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4):145–147.

Demoucron, M. (2008). *On the control of virtual violins-Physical modelling and control of bowed string instruments*. PhD thesis, Universite Pierre et Marie Curie-Paris VI; Royal Institute of Technology, Stockholm.

Devaney, J. and Ellis, D. P. W. (2008). An Empirical Approach to Studying Intonation Tendencies in Polyphonic Vocal Performances. *Journal of Interdisciplinary Music Studies*, 2(1):141–156.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, 30(9):1647–1669.

Dixon, S. (2005). An on-line time warping algorithm for tracking musical performances. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 1727–1728.

Dixon, S., Goebl, W., and Widmer, G. (2002). The performance worm: real time visualization of expressed based on Langner's tempo-loudness animation. In *International Computer Music Conference*.

Donald, E. (2012). Private communication.

Fletcher, H. and Sanders, L. C. (1967). Quality of violin vibrato tones. *The Journal of the Acoustical Society of America*, 41(6):1534–1544.

Fyk, J. (1993). Static and dynamic model of musical intonation. In *SMAC 93. Proceedings of the Stockholm Music Acoustics Conference*, pages 89–95.

Fyk, J. (1995). *Melodic intonation, psychoacoustics and the violin*. Poland:Organon publishing house.

Gabrielsson, A. (1987). Once again: the theme from Mozart's piano Sonata in A Major (k. 331). *Action and perception in rhythm and music*, 55:81–103.

Gabrielsson, A. (1999). The Performance of Music. In Deutsch, D., editor, *The Psychology of Music*, volume 2, pages 501–602. San Diego: Academic Press, 2nd edition.

Gabrielsson, A. (2003). Music Performance Research at the Millennium. *Psychology of Music*, 31(3):221–272.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, 40(10):1206–1220.

Glowinski, D., Coletta, P., Volpe, G., Camurri, A., Chiorri, C., and Schenone, A. (2010). Multi-Scale Entropy Analysis of Dominance in Social Creative Activities. *Complexity*, pages 1035–1038.

Glowinski, D., Mancini, M., Cowie, R., Camurri, A., Chiorri, C., and Doherty, C. (2013). The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves. *Frontiers in psychology*, 4:841.

Glowinski, D., Torres-Eliard, K., Chiorri, C., Camurri, A., and Grandjean, D. (2012). Can naive observers distinguish a violinist's solo from an ensemble performance? A pilot study. In *Third international workshop on social behaviour in music at ACM ICMI*.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *The Journal of the Acoustical Society of America*, 110(1):563–572.

Goebl, W., Dixon, S., De Poli, G., Friberg, A., Bresin, R., and Widmer, G. (2008). Sense in expressive music performance: Data acquisition, computational studies, and models. In Polotti, P. and Rocchesso, D., editors, *Sound to sense-sense to sound: A state of the art in sound and music computing*, pages 195–242. Logos Berlin.

Goebl, W. and Palmer, C. (2008). Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, 186(3):471–479.

Goebl, W. and Palmer, C. (2009). Synchronization of Timing and Motion Among Performing Musicians. *Music Perception*, 26(5):427–438.

Good, M. and Others (2001). MusicXML: An internet-friendly format for sheet music. In *XML Conference and Expo*, pages 3–4. Citeseer.

Goodman, E. (2002). *Musical Performance: Ensemble Performance*. Cambridge University Press, Cambridge.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438.

Guettler, K., Schoonderwaldt, E., and Askenfelt, A. (2003). Bow speed or bowing position: Which one influences spectrum the most?

Hajda, J. M., Kendall, R. A., Carterette, E. C., and Harshberger, M. L. (1997). Methodological issues in timbre research. In Deliege, I. and Sloboda, J. A., editors, *The Perception and Cognition of Music*, pages pp. 253–306. Psychology Press.

Heimann, M. (1958). *Exercises for String Quartet*. European String Teachers Association (Denmark branch), ACMP Chamber Music Network, 2007 edition.

Henderson, M. T. (1936). *Rhythmic organization in artistic piano performance.* PhD thesis, University of Iowa.

Honing, H. (2006). Evidence for tempo-specific timing in music using a web-based experimental setup. *Journal of Experimental Psychology; Human Perception and Performance*, 32(3):780–786.

Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary issues in communication science and disorders (CICSD)*, 31:122.

Kalin, G. (2005). *Formant frequency adjustment in barbershop quartet singing*. PhD thesis, KTH.

Keller, P. E. (2008). Joint action in music performance. In Morganti, F., Carassa, A., and Riva, G., editors, *Emerging communication*, volume 10, pages 205–221. IOS press.

Keller, P. E. and Appel, M. (2010). Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles. *Music Perception*, 28(1):27–46.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, pages 81–93.

Kirschner, S. and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31(5):354–364.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand. *Structure and perception of electroacoustic sound and music*, 9:43–53.

Krumhansl, C. L. and Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3):739.

Kulp, C. W. and Schlingmann, D. (2009). Using Mathematica to Compose Music and Analyze Music with Information Theory. In *Mathematics and Computation in Music*, pages 441–448. Springer.

Langner, J. and Goebl, W. (2003). Visualizing Expressive Performance in Tempo—Loudness Space. *Computer Music Journal*, 27(4):69–83.

Lehmann, A. C. and Kopiez, R. (2010). The difficulty of discerning between composed and improvised music. *Musicae Scientiae*, 14(2):113–129.

Loehr, J. D. and Palmer, C. (2009). Subdividing the beat: Auditory and motor contributions to synchronization. *Music Perception*, 26(5):415—-425.

Madrid, A. L. (2009). Why Music and Performance Studies? Why Now?: An Introduction to the Special Issue. *TRANS-Revista Transcultural de Música*, (13):1.

Maestre, E. (2009). *Modeling instrumental gestures: an analysis/synthesis framework for violin bowing*. PhD thesis, Universitat Pompeu Fabra.

Maestre, E., Blaauw, M., Bonada, J., Guaus, E., and Pérez, A. (2010). Statistical modeling of bowing control applied to violin sound synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(4):855–871.

Maestre, E., Bonada, J., Blaauw, M., Perez, A., and Guaus, E. (2007). Acquisition of violin instrumental gestures using a commercial EMF device. In *International Computer Music Conference*, pages 386–393. NI-BIT, San Francisco: ICMA.

Maestre, E., Scavone, G. P., and Smith III, J. O. (2013). Digital modeling of bridge driving-point admittance from measurements on violin-family instruments. In *Stockholm Music Acoustics Conference (SMAC 2013)*. Creative Commons License.

Marchini, M. (2014). *Analysis of Ensemble Expressive Performance in String Quartets: a Statistical and Machine Learning Approach*. phdthesis, Univesitat Pompeu Fabra.

Marchini, M., Papiotis, P., and Maestre, E. (2012). Timing synchronization in string quartet performance: a preliminary study. *International Workshop on Computer Music Modeling and Retrieval (CMMR12)*, pages 177–185.

Marchini, M., Papiotis, P., Pérez, A., and Maestre, E. (2011). A Hair Ribbon Deflection Model for Low-intrusiveness Measurement of Bow Force in Violin Performance. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 481–486.

Mason, J. A. (1960). Comparison of Solo and Ensemble Performances with Reference to Pythagorean, Just, and Equi-Tempered Intonations. *Journal of Research in Music Education*, 8(1):31.

Mattheson, J. (1739). *Der vollkommene Capellmeister*. Bärenreiter Verlag.

Mayor, O., Llimona, Q., Marchini, M., Papiotis, P., and Maestre, E. (2013). repoVizz: a framework for remote storage, browsing, annotation, and exchange of multi-modal data. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 415–416. ACM.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological research*, 58(3):177–192.

Meyer, J. (1993). Vibrato sounds in large halls. In *Stockholm Music Acoustics Conference*, pages 117–121.

Miles-Huber, D. (1991). The MIDI manual. *USA. Howard W. Sams*.

Moore, G. P. and Chen, J. (2010). Timings and interactions of skilled musicians. *Biological cybernetics*, 103(5):401–14.

Mozart, L. (1756). *Versuch einer gründlichen Violinschule: entworfen und mit 4. Kupfertafeln sammt einer Tabelle versehen*. HL Grahl.

Müller, M. (2015). *Fundamentals of Music Processing*. Springer Science.

Navarro, G. (2011). Private communication.

Nickerson, J. F. (1949). Intonation of Solo and Ensemble Performance of the Same Melody. *The Journal of the Acoustical Society of America*, 21(6):593.

Niedermayer, B. (2012). *Accurate Audio-to-Score Alignment-Data Acquisition in the Context of Computational Musicology*. PhD thesis.

Nieto, O. (2015). Discovering Structure in Music: Automatic Approaches and Perceptual Evaluations.

Orio, N. and Déechelle, F. (2001). Score Following Using Spectral Analysis and Hidden Markov Models. In *Proc. of the International Computer Music Conference, La Habana, CU*, volume 4, pages 125–129.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of experimental psychology: human perception and performance*, 15(2):331.

Palmer, C. (1997). Music Performance. *Annual Review of Psychology*, 48(1):115–138.

Palmer, C. (2013). Music Performance: Movement and Coordination. *The Psychology of Music*, pages 405–422.

Palmer, C., Koopmans, E., Loehr, J. D., and Carter, C. (2009). Movement-related feedback and temporal accuracy in clarinet performance. *Music Perception*, 26(5):439—-449.

Pantic, M., Cowie, R., D'Errico, F., Heylen, D., Mehu, M., Pelachaud, C., Poggi, I., Schroeder, M., and Vinciarelli, A. (2011). Social Signal Processing: The Research Agenda. In Moeslund, T. B., Hilton, A., Krüger, V., and Sigal, L., editors, *Visual Analysis of Humans*, pages 511–538. Springer London.

Papich, G. and Rainbow, E. (1974). A pilot study of performance practices of twentieth-century musicians. *Journal of Research in Music Education*, 22(1):24–34.

Papiotis, P., Herrera, P., and Marchini, M. (2013a). Aural-Based Detection and Assessment of Real Versus Artificially Synchronized String Quartet Performance. *Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyv{ä}skyl{ä}, Finland, 11th-15th June 2013. Geoff Luck & Olivier Brabant (Eds.). ISBN 978-951-39-5250-1*.

Papiotis, P., Maestre, E., Marchini, M., and Perez Carrillo, A. (2011). Synchronization of intonation adjustments in violin duets: towards an objective evaluation of musical interaction. In *14th International Conference on Digital Audio Effects (DAFx11)*, pages 417–423.

Papiotis, P., Marchini, M., and Maestre, E. (2012a). Computational analysis of solo versus ensemble performance in string quartets: dynamics and intonation. In Cambouropoulos, E., Tsougras, C., Mavromatis, P., and Pastiades, K., editors, *12th International Conference of Music Perception and Cognition (ICMPC12)*, Thessaloniki.

Papiotis, P., Marchini, M., and Maestre, E. (2013b). Multidimensional analysis of interdependence in a string quartet. In *International Symposium on Performance Science (ISPS2013)*, Vienna, Austria.

Papiotis, P., Marchini, M., Maestre, E., and Perez, A. (2012b). Measuring ensemble synchrony through violin performance parameters: a preliminary progress report. In *Intelligent Technologies for Interactive Entertainment*, pages 267–272. Springer Berlin Heidelberg.

Papiotis, P., Marchini, M., Perez-Carrillo, A., and Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. *Frontiers in Psychology*, 5(September):963.

Parncutt, R. and McPherson, G. (2002). *The science and psychology of music performance: Creative strategies for teaching and learning*. Oxford University Press.

Patel, A. D. (2010). *Music, language, and the brain*. Oxford university press.

Pecenka, N. and Keller, P. E. (2009). Auditory pitch imagery and its relationship to musical synchronization. *Annals of the New York Academy of Sciences*, 1169:282–6.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., and McAdams, S. (2011). The Timbre Toolbox: extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5):2902–16.

Peeters, G., McAdams, S., and Herrera, P. (2000). Instrument sound description in the context of MPEG-7. In *Proceedings of the 2000 International Computer Music Conference*, pages 166–169. Citeseer.

Pereda, E., Quiroga, R. Q., and Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, 77(1-2):1–37.

Perez Carrillo, A. (2009). *Enhacing spectral sintesis techniques with performance gestures using the violin as a case study*. Phd thesis, Universitat Pompeu Fabra.

Perez Carrillo, A., Bonada, J., Maestre, E., Guaus, E., and Blaauw, M. (2007). Combining performance actions with spectral models for violin sound transformation. *Proceedings of 19th International Congress on Acoustics*.

Pressing, J. (1988). Improvisation: methods and models. *Generative processes in music: The psychology of performance, improvisation, and composition*, pages 129–178.

Quantz, J. J. (1752). *Versuch einer Anweisung die Flöte traversiere zu spielen*.

Rasamimanana, N. H., Flety, E., and Bevilacqua, F. (2006). Gesture Analysis of Violin Bow Strokes. In *Lecture Notes in Computer Science*, volume 3881, pages 145–155. Springer Verlag.

Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In Sloboda, J., editor, *Generative processes in music: The psychology of performance, improvisation, and composition*, pages 70–90. Oxford: Clarendon Press.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Traumerei. *The Journal of the Acoustical Society of America*, 92(5):2546–2568.

Repp, B. H. (1995). Expressive timing in Schumann's Traumerei: An analysis of performances by graduate student pianists. *The Journal of the Acoustical Society of America*, 98(5):2413–2427.

Repp, B. H. (1996). The dynamics of expressive piano performance: Schumann's "Tr{ä}umerei"revisited. *The Journal of the Acoustical Society of America*, 100(1):641–650.

Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception*, pages 409–438.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6):969–992.

Repp, B. H., Keller, P. E., and Jacoby, N. (2012). Quantifying phase correction in sensorimotor synchronization: empirical comparison of three paradigms. *Acta psychologica*, 139(2):281–90.

Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychonomic bulletin & review*, 20(3):403–52.

Repp, B. H., Windsor, L., and Desain, P. (2002). Effects of Tempo on the Timing of Simple Musical Rhythms. *Music Perception*, 19(4):565–593.

Rink, J. (2002). *Musical performance: a guide to understanding*. Cambridge University Press.

Rink, J. (2003). In respect of performance: The view from Musicology. *Psychology of Music*, 31(3):303–323.

Salamon, J. and Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(6):1759–1770.

Schelleng, J. C. (1973). The bowed string and the player. *The Journal of the Acoustical Society of America*, 53(1):26.

Schögler, B. (2000). Studying temporal co-ordination in jazz duets. *Musicae Scientiae*, 3(1 suppl):75–91.

Schoonderwaldt, E. and Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *The Journal of the Acoustical Society of America*, 126(5):2695–2708.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2):461.

Schreiber, T. and Schmitz, A. (2000). Surrogate time series. *Physica D: Nonlinear Phenomena*, 142(3-4):346–382.

Seashore, C. E. (1936). *Objective analysis of musical performance*, volume 4. The University Press.

Seashore, C. E. (1938). *Psychology of music*. Courier Corporation.

Seashore, H. G. (1932). The hearing of the pitch and intensity in vibrato. In Seashore, C. E., editor, *University of Iowa studies in the psychology of music: Vol I. The vibrato*, pages 213–235. Iowa City: University of Iowa.

Serra, X., Magas, M., Benetos, E., Chudy, M., Dixon, S., and Flexer, A. (2013). *Roadmap for Music Information ReSearch*.

Shackford, C. (1961). Some aspects of perception. I: Sizes of harmonic intervals in performance. *Journal of Music Theory*, pages 162–202.

Shaffer, L. H. (1980). Analysing Piano Performance: A Study of Concert Pianists. *Advances in Psychology*, 1:443–455.

Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101.

Sundberg, J., Friberg, A., and Frydén, L. (1989). Rules for Automated Performance of Ensemble Music. *Contemporary Music Review*, 3.

Tinker, B. (2006). Harmonic analyzer.

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, 91(6):3540–3550.

Vincent, E., Bertin, N., Gribonval, R., and Bimbot, F. (2014). From Blind to Guided Audio Source Separation: How models and side information can improve the separation of sound. *Signal Processing Magazine, IEEE*, 31(3):107–115.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In *Gesture and sign language in human-computer interaction*, pages 241–253. Springer.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34(1):97–113.

Widmer, G., Dixon, S., Goebl, W., Pampalk, E., and Tobudic, A. (2003). In Search of the Horowitz Factor. *AI Magazine*, pages 111–130.

Widmer, G. and Goebl, W. (2004). Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research*, 33(3):203–216.

Williamon, A. and Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, 6(1):53–72.

Wiltermuth, S. S. and Heath, C. (2009). Synchrony and Cooperation. *Psychological Science*, 20(1):1–5.

Wing, A. M., Endo, S., Bradbury, A., and Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, 11(93):20131125–20131125.

Young, V. M. and Colman, A. M. (1979). Some psychological processes in string quartets. *Psychology of Music*, 7(1):12–18.

Zapata, J. (2013). *Comparative evaluation and combination of automatic rhythm description systems*. PhD thesis, Universitat Pompeu Fabra, Barcelona.

Appendices

Appendix A

SCORES OF STRING QUARTET EXERCISES RECORDED IN THE MAIN EXPERIMENT

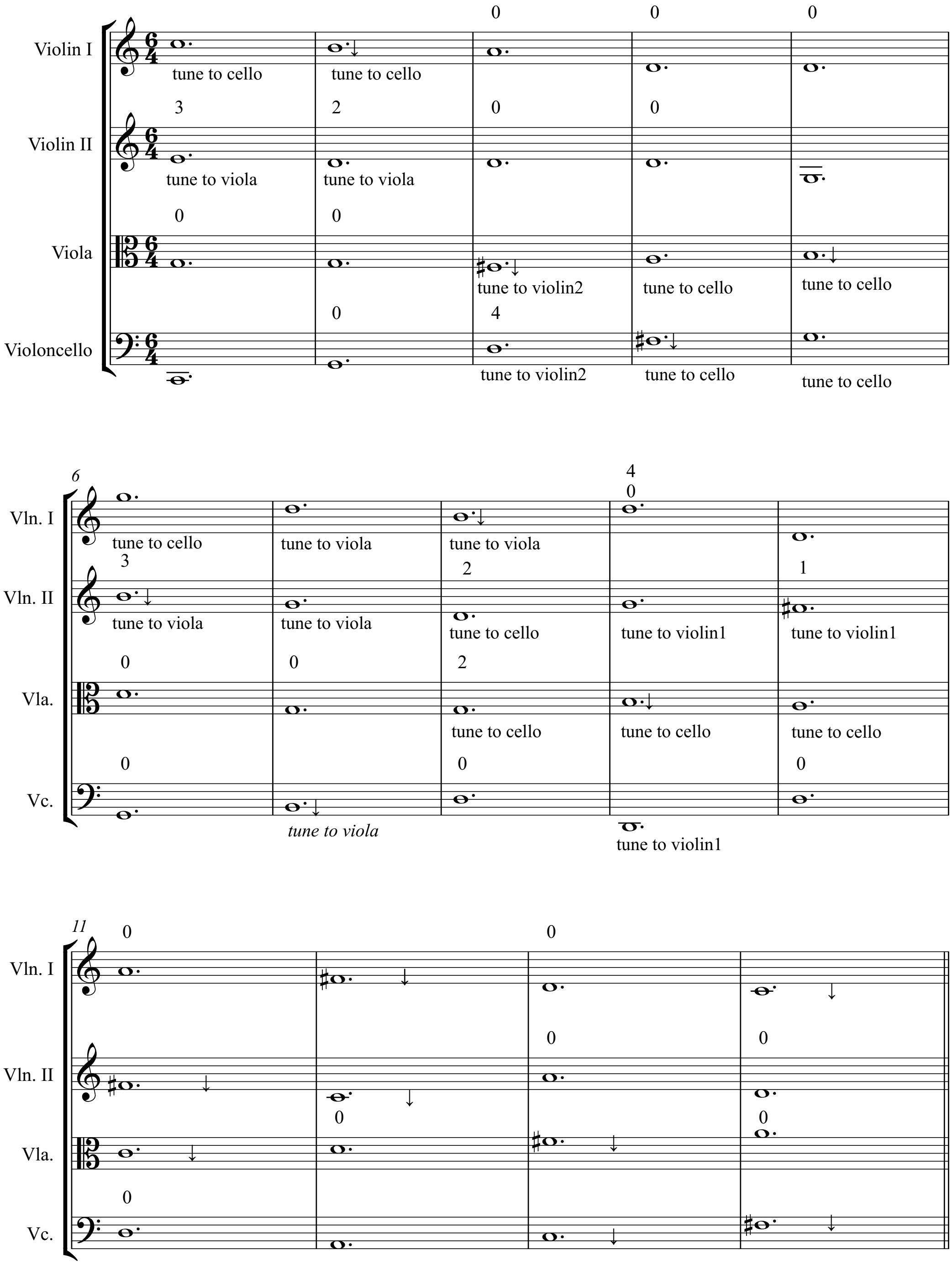
In this appendix we provide the musical scores for the exercises recorded as part of the main experiment detailed in Chapter 3.

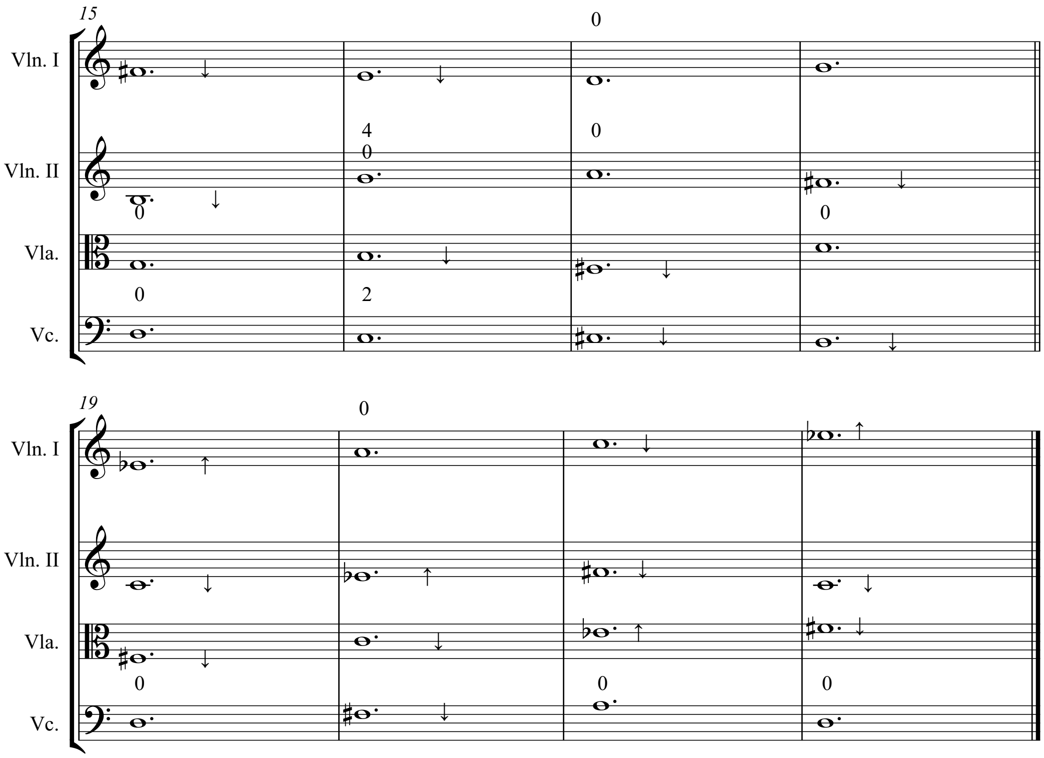
All of the following exercises are based on an exercise handbook by Heimann (1958); it is provided by Associated Chamber Music Players (ACMP) and the Danish branch of the European String Teachers Association (ESTA) , and can be found online in its entirety at http://www.acmp.net/publications.

Intonation exercise

I1

Choral intonation, correct immediately



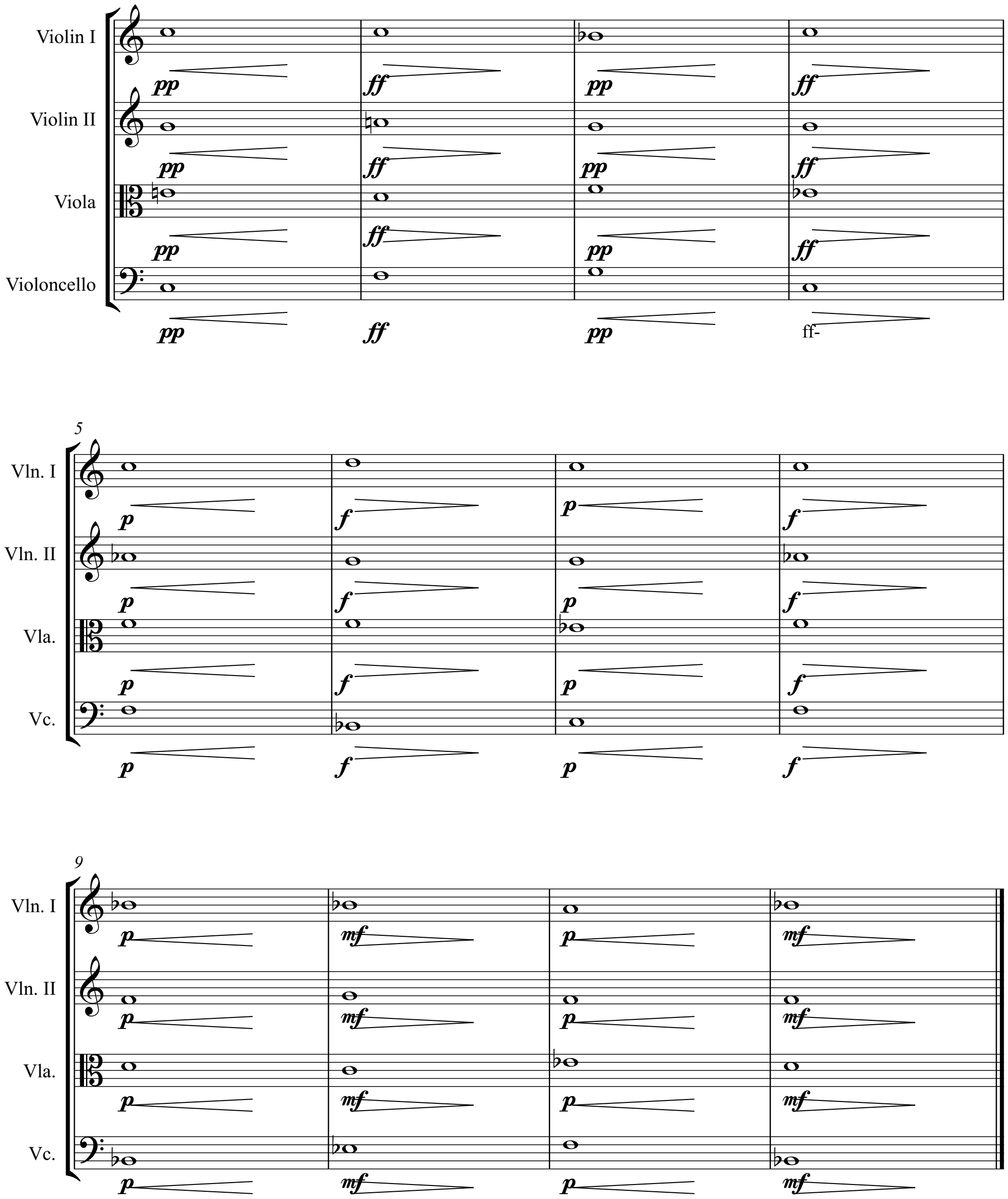

15
Vln. I
Vln. II
Vla.
Vc.
19
Vln. I
Vln. II
Vla.
Vc.

Dynamics exercises

D1

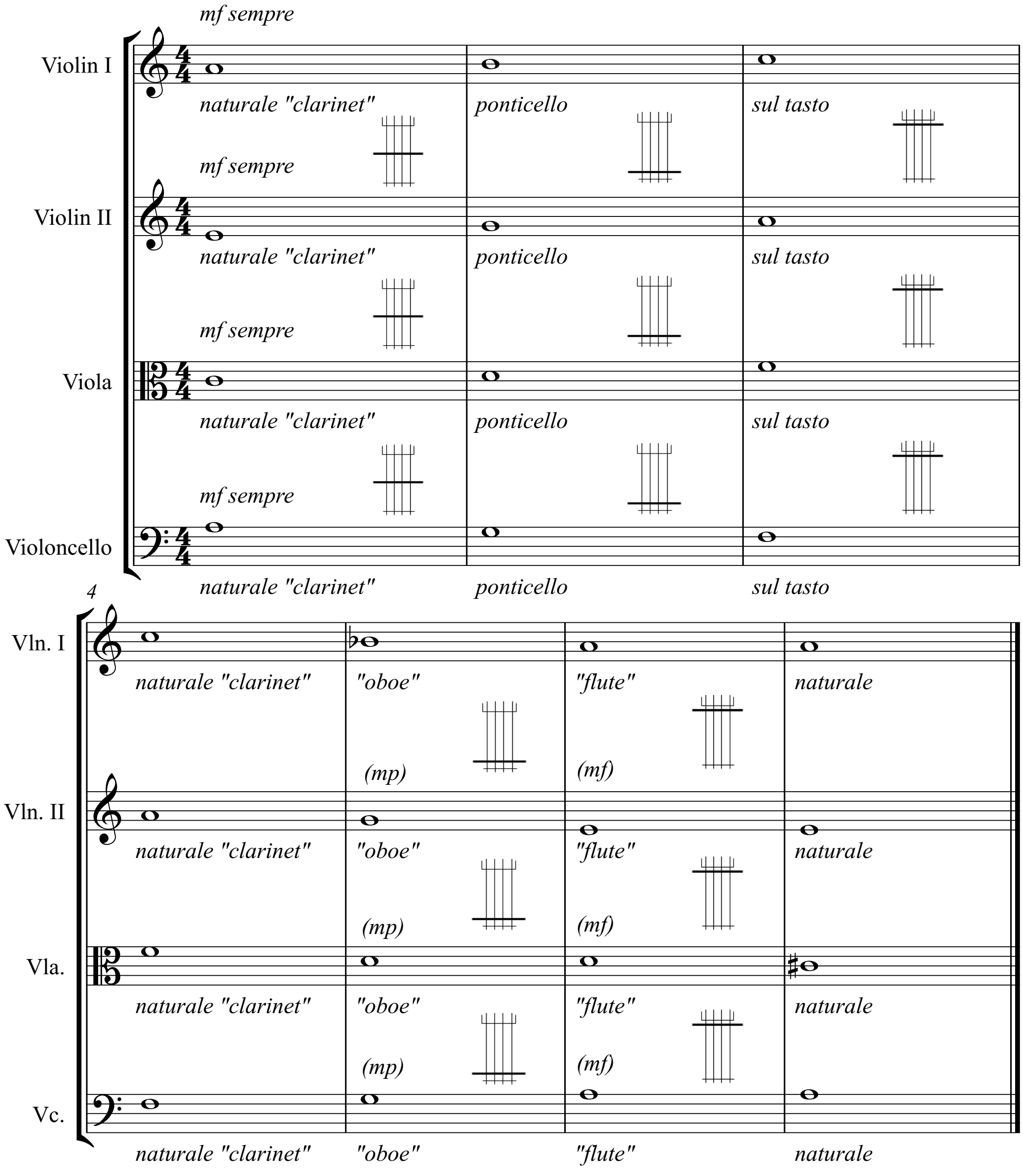
Distinct contrasts,
sustain the respective dynamic shading unaltered until the next note.

D2


Violin I
Violin II
Viola
Violoncello
pp
ff
ff-
5
Vln. I
Vln. II
Vla.
Vc.
p
f
9
mf

Timbre exercise

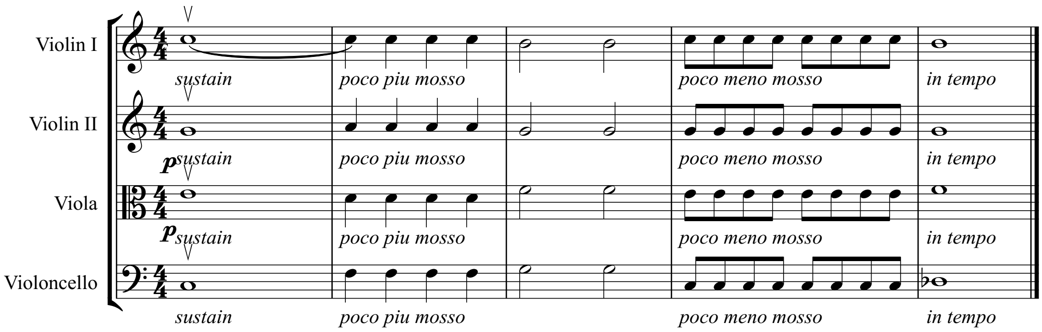
T1



Rhythm exercises

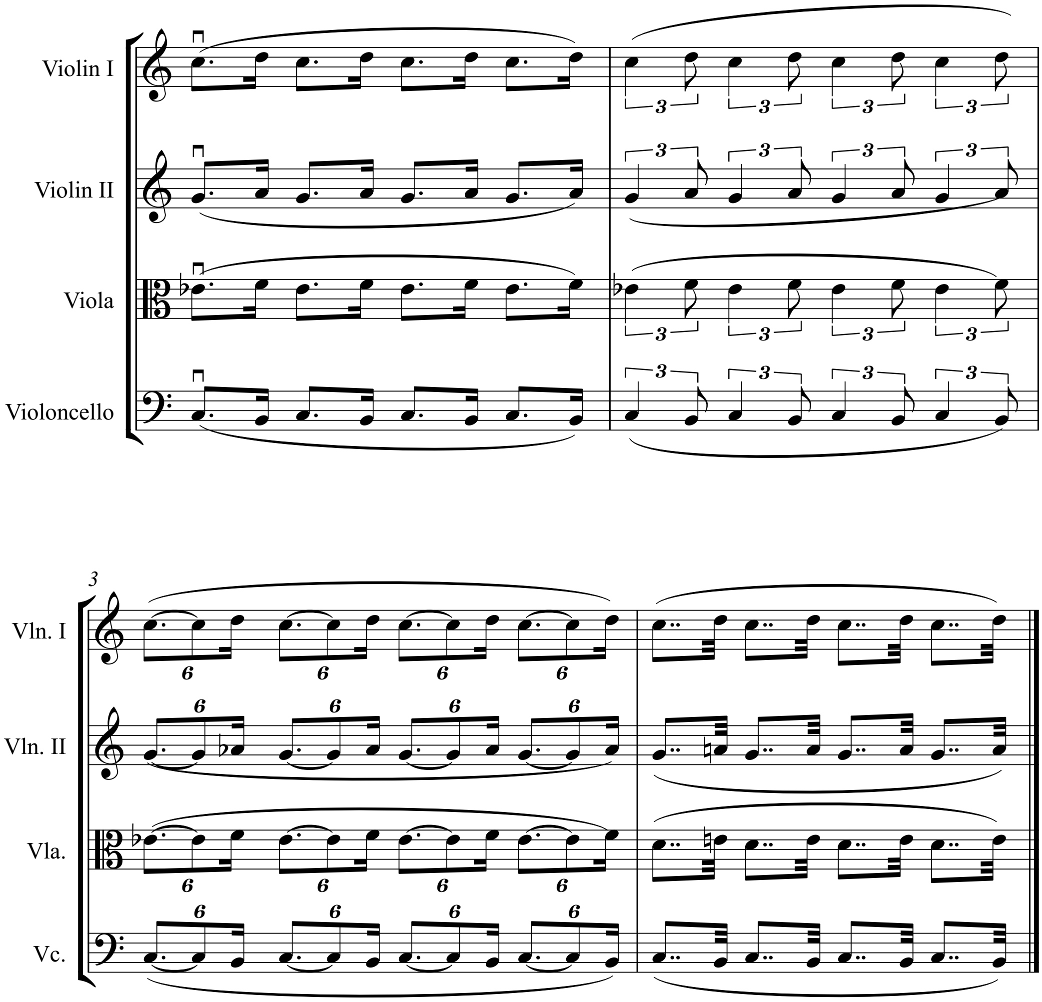
R1

Repeat 4 times



R2

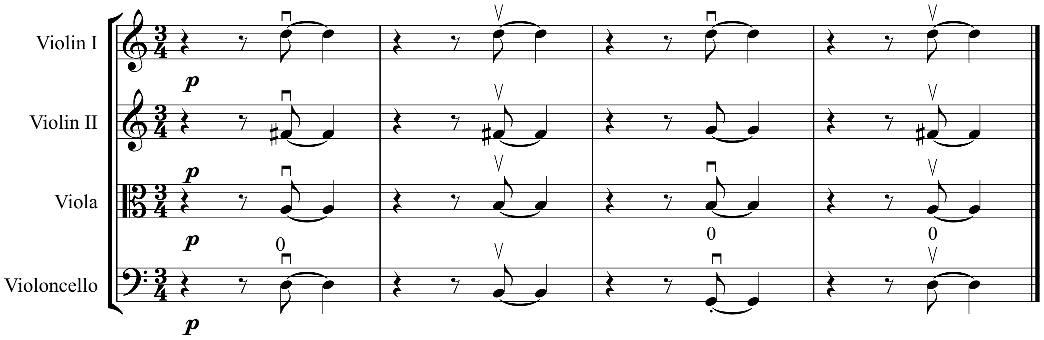
Clear distinction between patterns. Repeat 4 times.



Unity of Execution exercises

U1

Attack, change, relieve and finish exactly together. Repeat 4 times.

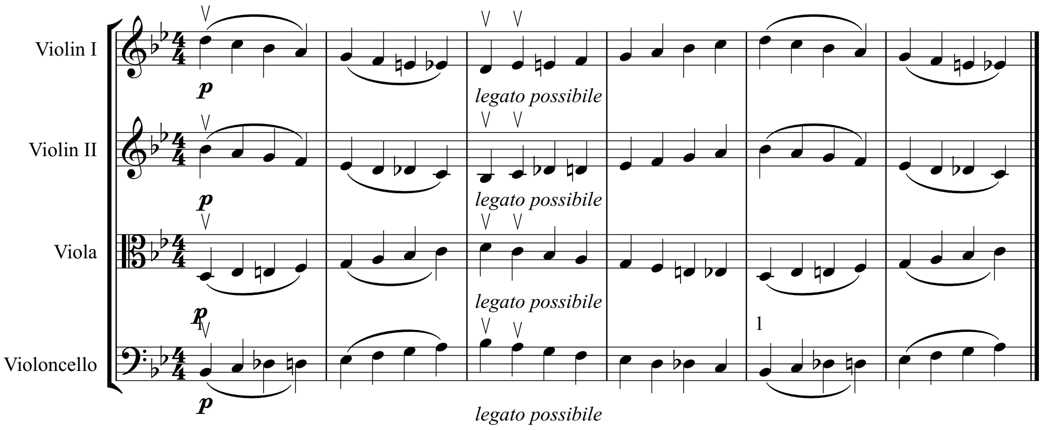


U2

Tranquillo; no break, should be played as one instrument. Repeat 4 times.

Phrasing exercise

P1



Appendix B

QUESTIONNAIRES

In this appendix we provide the questionnaire forms that were filled out by the musicians following the recording of each exercise, as well as a preliminary analysis of the questionnaire results.

B.1. Questionnaire form

Rate the difficulty of your part (1-very easy, 5-very difficult)

1	2	3	4	5
☐	☐	☐	☐	☐

Rate the difficulty of the exercise, as an ensemble task (1-very easy, 5-very difficult)

1	2	3	4	5
☐	☐	☐	☐	☐

Rate how succesfully you performed your part (1-not at all, 5-very succesfully)

1	2	3	4	5
☐	☐	☐	☐	☐

Rate how succesfully the ensemble performed the task (1-not at all, 5-very succesfully)

1	2	3	4	5
☐	☐	☐	☐	☐

How clear was the goal of the exercise? (1-not clear at all, 5-very clear)

1	2	3	4	5
☐	☐	☐	☐	☐

How important was your part in achieving the goal of the exercise? (1-not important, 3-very important)

1	2	3
☐	☐	☐

Were you more focused on playing your part or on achieving ensemble cohesion? (1-completely on playing my part, 5-completely on achieving cohesion)

1	2	3	4	5
☐	☐	☐	☐	☐

Do you think there was a 'leader' role?

Yes	No
☐	☐

If yes, which instrument?

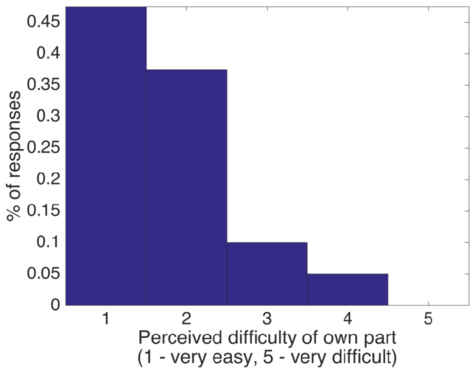
violin1	violin2	viola	cello
☐	☐	☐	☐

B.2. Analysis of questionnaire data

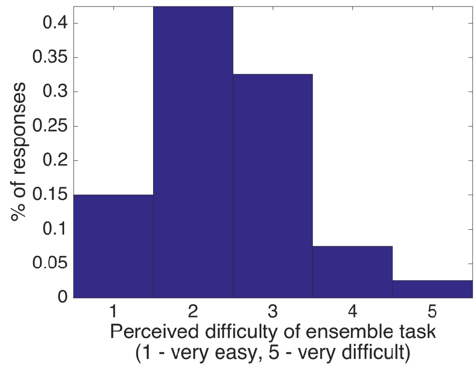
We present questionnaire results for all of the exercises recorded in our main experiment (two exercises on dynamics, one exercise on intonation, one exercise on timbre, two exercises on rhythm, one exercise on phrasing, and two exercises on unity of execution); each questionnaire was filled out following the ensemble recording of each exercise.

B.2.1. Difficulty

Regarding the perceived difficulty of each individual part of the exercise as well as the perceived difficulty of the exercise as an ensemble task, the results can be seen in Figures B.1a and B.1b, respectively.



(a) in terms of the individual parts of each performer



(b) as an ensemble task

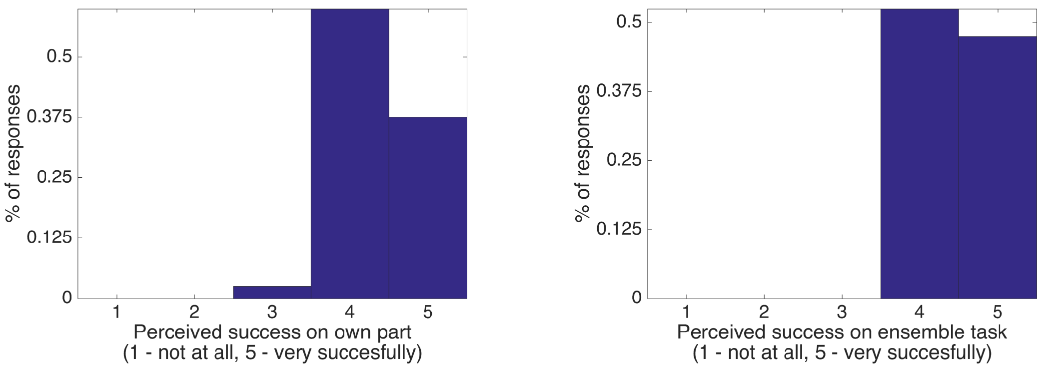
Figure B.1: Questionnaire results for perceived difficulty of the performed exercises.

It can be seen that the purported simplicity of the exercises as individual tasks is indeed true; in 85% of the responses, the subjects rated their own task the exercise as “very easy” or ”easy”, while only on two occasions was the task rated as “moderately difficult” (exercise R2 for the violist, and exercise T1 for violinist 1).

At the same time, it is also seen that the overall perceived difficulty of each exercise as an ensemble task is increased compared to the perceived difficulty as an individual task, which is also in line with the rationale of using these exercises as material to work on ensemble interdependence.

B.2.2. Degree of success

The next question deals with the degree to which the performers believed to have successfully performed their own part, as well as the shared goal of the exercise; the results can be seen in Figures B.2a and B.2b, respectively.



(a) in terms of the individual parts of each performer

(b) in terms of the ensemble goal

Figure B.2: Questionnaire results for perceived degree of success with which the exercises were performed.

Both in regard to individual as well as the ensemble performance, it can be seen that the overwhelming majority of ratings are between "moderately successful" and "very successful" performance of the individual task and the shared goal of the ensemble, respectively.

We additionally computed the correlation between perceived difficulty and perceived success of the performance. Regarding the relationship between individual task difficulty and individual task success, we found a negative correlation ($r = -0.52, p < 0.05$) between the variables - signifying that the easier individual tasks are related to more successful individual performance; a similar result with a somewhat weaker correlation ($r = -0.35, p < 0.05$) was also found between individual task difficulty and ensemble task success. In the same vein, the difficulty of the ensemble goal was also found to be negatively correlated with the success of both the individual task ($r = -0.44, p < 0.05$) as well as the ensemble goal ($r = -0.52, p < 0.05$). Finally, we found a positive correlation between the success of the individual task and the success of the ensemble task ($r = 0.51, p < 0.05$), as well as the difficulty of the individual task and the difficulty of the ensemble task ($r = 0.70, p < 0.05$).

Given the way difficulty and success ratings are mostly distributed around extreme values (very easy/moderately easy for difficulty and moderately successfully/very successfully for the achievement of the shared goal), the above results are not surprising to see. They do however demonstrate the influence of the underlying score on both the difficulty and success of the shared goal of the ensemble.

B.2.3. Clarity of the shared goal

There is little to report regarding the clarity of the shared goal; for all exercises and performers, the goal of the exercise was rated as "very clear" with only two exercises being rated 'moderately clear' by a subset of the ensemble; the intonation exercise (I1) with added annotations (by violin 2 and cello) and the timbre exercise (T1) (by violin 2 and viola). The fact that both exercises featured annotations not typically seen in musical scores (see Appendix A) could potentially explain this result.

B.2.4. Perceived importance of each performer's individual part

In this question, the performers were tasked with judging the importance of their own part in achieving the shared goal of the ensemble. Violinist 1 consistently marked her individual part in every exercise as "very important", while the cellist consistently marked his individual part as of "medium importance". Violinist 2 marked her individual part as "very important" for all exercises but exercise R1, where her part was marked as of "medium importance", while the violist marked her part as of "medium importance" for all exercises but D1, I1 (without annotations) and U1.

A one-way ANOVA was carried out to assess whether the differences between the performers in terms of perceived importance were statistically significant. The obtained results ($F = 27.6, p < 0.001$) in conjunction with a post hoc test showed that mean perceived importance for violinist 1 was significantly different from violist and cellist, and the same result was seen for violinist 2 (significantly different from the violist and cellist). No significant differences were found between violinists 1 and 2, or violist and cellist.

The above results demonstrate a difference between the performers, namely that violinists 1 and 2 tend to consider their contribution highly important to ensemble cohesion, while the other two performers adopted a more moderate view of their own perceived importance. Whether this is objectively assessed (and thus purely related to the role in the ensemble) or dependent on the predisposition of the performers remains to be seen in a future study.

B.2.5. Direction of focus

In this question, the performers were asked to report whether they were focusing more on playing their parts or achieving ensemble cohesion. The results can be seen in Figure B.3.

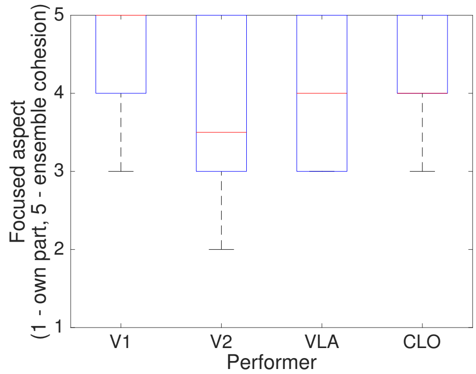


Figure B.3: Questionnaire results for the direction of focus of each performer.

All performers appear to be focusing mostly on achieving ensemble cohesion. Violinist 2 and violist show a slight tendency towards balanced focus between ensemble cohesion and the individual task, although a one-way ANOVA did not reveal any significant differences between the mean ratings of the performers.

B.2.6. Perceived leadership

This final two-part question dealt with the existence of a "leader role" among the performers. The existence of a leader was reported for 70% of the ratings; Figure B.4 shows the distribution for the performer perceived as a leader:

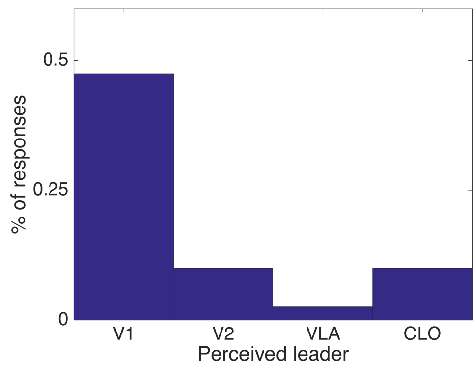


Figure B.4: Questionnaire results for perceived existence of a leader role within the quartet.

In nearly 50% of the cases where a leader role was thought to exist, that role was filled by violinist 1. Viola was considered the leader only once, while violinist 2 and cello were unilaterally considered to hold a leader role in one exercise each (R1 for violinist 2 and U1 for the cellist).

Figure B.5 additionally shows the tendency of each performer to report the existence of a leader, while Figure B.6 shows detailed results regarding the response of each musician for each exercise.

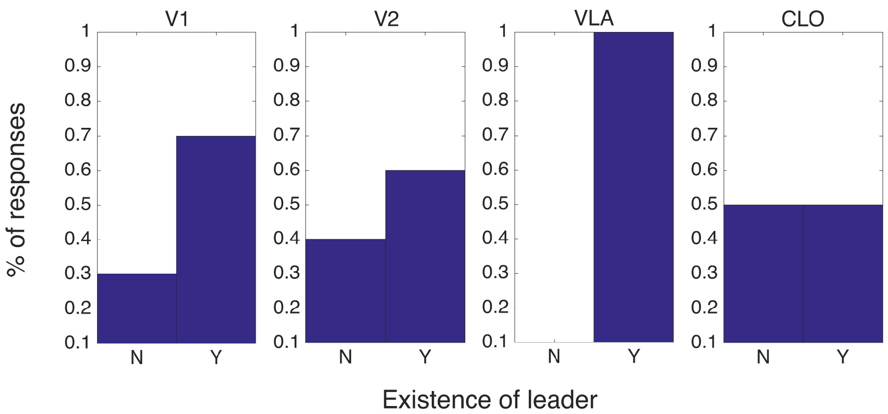


Figure B.5: Questionnaire results for perceived existence of a leader role within the quartet, for each performer individually.

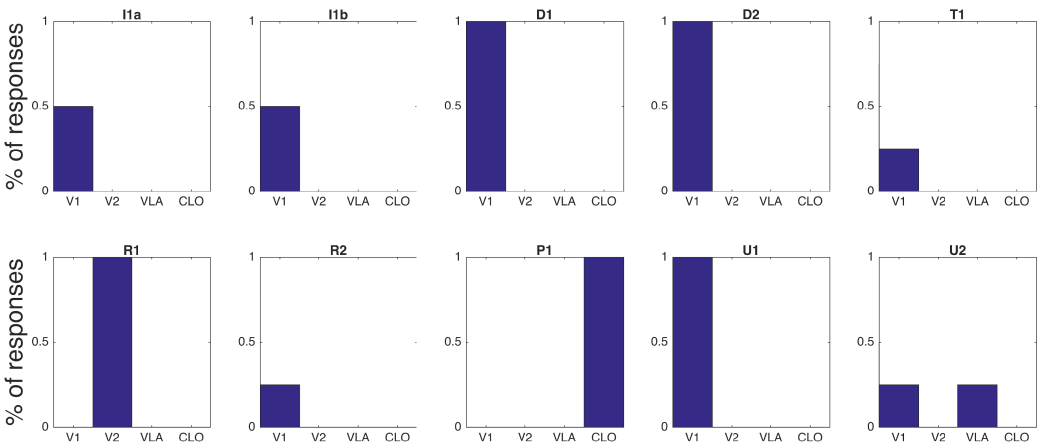


Figure B.6: Questionnaire results for perceived existence of a leader role within the quartet, for each exercise.

More than any other potential conclusion, what is seen in the above two figures is that each member of the ensemble has different views on the existence of a leader, and that this view changes with the exercise; for example, half of the exercises

(D1, D2, R1, P1, U1) demonstrate a unanimous decision on the existence of a specific leader, which itself varies from violinist 1, 2 or the cellist; on the other hand, the rest of the exercises demonstrate a divergence of opinion both regarding the existence of a leader or not, as well as which performer fills that role.

B.2.7. Discussion

The collected responses suggest that the ensemble encountered little difficulty in performing the exercises they were given, both in terms of the individual part as well as in terms of the shared goal of each exercise; it was also seen that achieving the shared goal was generally considered as more difficult than the individual part, giving weight to the claim that the main challenge of the recorded exercises lies in achieving interdependent performance. Additionally, the shared goal was consistently rated as "very clear" in virtually all cases.

As expected, the perceived difficulty of the exercise is negatively correlated to the perceived success with which it was performed, while successful performance of each musician's individual part appears positively correlated with successful performance in terms of the shared goal. While these results are within ordinary expectations, they do highlight the link between performed material and potential for interdependence, as well as provide some hints on the optimal conditions for its establishment.

Finally, we were able to gain some insights on the profile of the performers regarding their self-assessment of importance in achieving the shared goal, as well as their tendency to consider the existence of a leader in terms of interdependence. Regarding the self-assessed importance we were able to find two statistically significant groups with divergent assessments: violinist 1 and violinist 2 tended to consider their individual parts as highly important for the achievement of the shared goal, while the violist and cellist tended to have a more moderate self-assessment. Regarding leadership, the results suggest that each of the performed exercises resulted in different assessments regarding leadership, while the perceived leader could vary from exercise to exercise as well.

This study has presented an overview of the responses and a few clear tendencies among the ensemble, although there are still several questions to be explored from the questionnaire data as well as from their combination with the quantitative results presented in Section 4.2; an in-depth examination of the results per exercise and performer could serve as a starting point.

Appendix C

INDEX OF RECORDINGS

Short title	Long title	Duration (s)
I1S1	Intonation Exercise 1 Solo violin 1	143
I1S2	Intonation Exercise 1 Solo violin 2	124
I1S3	Intonation Exercise 1 Solo viola	126
I1S4	Intonation Exercise 1 Solo cello	133
I1E1	Intonation Exercise 1 Ensemble (no annotations)	168
I1E2	Intonation Exercise 1 Ensemble (with annotations)	144
IP1S1	Intonation Piece excerpt 1 Solo violin 1	16
IP1S2	Intonation Piece excerpt 1 Solo violin 2	15
IP1S3	Intonation Piece excerpt 1 Solo viola	16
IP1S4	Intonation Piece excerpt 1 Solo cello	17
IP1E	Intonation Piece excerpt 1 Ensemble	43
IP2S1/DP2S1	Intonation / Dynamics Piece excerpt 2 Solo violin 1	16
IP2S2/IP2S2	Intonation / Dynamics Piece excerpt 2 Solo violin 2	15
IP2S3/DP2S3	Intonation / Dynamics Piece excerpt 2 Solo viola	16
IP2S4/DP2S4	Intonation / Dynamics Piece excerpt 2 Solo cello	17
IP2E/DP2E	Intonation / Dynamics Piece excerpt 2 Ensemble	43
D1S1	Dynamics Exercise 1 Solo violin 1	61
D1S2	Dynamics Exercise 1 Solo violin 2	51
D1S3	Dynamics Exercise 1 Solo viola	60
D1S4	Dynamics Exercise 1 Solo cello	59
D1E	Dynamics Exercise 1 Ensemble	63
D2S1	Dynamics Exercise 2 Solo violin 1	53
D2S2	Dynamics Exercise 2 Solo violin 2	41

D2S3	Dynamics Exercise 2 Solo viola	52
D2S4	Dynamics Exercise 2 Solo cello	49
D2E	Dynamics Exercise 2 Ensemble	53
DP1S1	Dynamics Piece excerpt 1 Solo violin 1	16
DP1S2	Dynamics Piece excerpt 1 Solo violin 2	15
DP1S3	Dynamics Piece excerpt 1 Solo viola	16
DP1S4	Dynamics Piece excerpt 1 Solo cello	17
DP1E	Dynamics Piece excerpt 1 Ensemble	43
T1S1	Timbre exercise Solo violin 1	57
T1S2	Timbre exercise Solo violin 2	51
T1S3	Timbre exercise Solo viola	60
T1S4	Timbre exercise Solo cello	55
T1E	Timbre exercise Ensemble	65
R1S1	Rhythm exercise 1 Solo violin 1	58
R1S2	Rhythm exercise 1 Solo violin 2	49
R1S3	Rhythm exercise 1 Solo viola	64
R1S4	Rhythm exercise 1 Solo cello	56
R1E	Rhythm exercise 1 Ensemble	59
R2S1	Rhythm exercise 2 Solo violin 1	62
R2S2	Rhythm exercise 2 Solo violin 2	49
R2S3	Rhythm exercise 2 Solo viola	64
R2S4	Rhythm exercise 2 Solo cello	56
R2E	Rhythm exercise 2 Ensemble	73
RP1S1	Rhythm excerpt 1 Solo violin 1	46
RP1S2	Rhythm excerpt 1 Solo violin 2	42
RP1S3	Rhythm excerpt 1 Solo viola	44
RP1S4	Rhythm excerpt 1 Solo cello	39
RP1E	Rhythm excerpt 1 Ensemble	42
RP2S1	Rhythm excerpt 2 Solo violin 1	28
RP2S2	Rhythm excerpt 2 Solo violin 2	38
RP2S3	Rhythm excerpt 2 Solo viola	31
RP2S4	Rhythm excerpt 2 Solo cello	24
RP2E	Rhythm excerpt 2 Ensemble	25
U1S1	Unity of Execution exercise 1 Solo violin 1	46
U1S2	Unity of Execution exercise 1 Solo violin 2	43
U1S3	Unity of Execution exercise 1 Solo viola	55
U1S4	Unity of Execution exercise 1 Solo cello	50
U1E	Unity of Execution exercise 1 Ensemble	45

U2S1	Unity of Execution exercise 2 Solo violin 1	63
U2S2	Unity of Execution exercise 2 Solo violin 2	61
U2S3	Unity of Execution exercise 2 Solo viola	64
U2S4	Unity of Execution exercise 2 Solo cello	63
U2E	Unity of Execution exercise 2 Ensemble	72
UP1S1	Unity of Execution excerpt 1 Solo violin 1	27
UP1S2	Unity of Execution excerpt 1 Solo violin 2	29
UP1S3	Unity of Execution excerpt 1 Solo viola	30
UP1S4	Unity of Execution excerpt 1 Solo cello	34
UP1E	Unity of Execution excerpt 1 Ensemble	38
UP2S1	Unity of Execution excerpt 2 Solo violin 1	43
UP2S2	Unity of Execution excerpt 2 Solo violin 2	41
UP2S3	Unity of Execution excerpt 2 Solo viola	42
UP2S4	Unity of Execution excerpt 2 Solo cello	40
UP2E	Unity of Execution excerpt 2 Ensemble	43
UP3S1/DP3S1	Unity of Execution/Dynamics excerpt 3 Solo violin 1	44
UP3S2/DP3S2	Unity of Execution/Dynamics excerpt 3 Solo violin 2	41
UP3S3/DP3S3	Unity of Execution/Dynamics excerpt 3 Solo viola	42
UP3S4/DP3S4	Unity of Execution/Dynamics excerpt 3 Solo cello	40
UP3E/DP3E	Unity of Execution/Dynamics excerpt 3 Ensemble	42
P1S1	Phrasing exercise 1 Solo violin 1	58
P1S2	Phrasing exercise 1 Solo violin 2	58
P1S3	Phrasing exercise 1 Solo viola	75
P1S4	Phrasing exercise 1 Solo cello	67
P1E	Phrasing exercise 1 Ensemble	63
PP1S1	Phrasing excerpt 1 Solo violin 1	46
PP1S2	Phrasing excerpt 1 Solo violin 2	57
PP1S3	Phrasing excerpt 1 Solo viola	56
PP1S4	Phrasing excerpt 1 Solo cello	54
PP1E	Phrasing excerpt 1 Ensemble	58
PP2S1/RP3S1	Phrasing excerpt 2 / Rhythm excerpt 3 Solo violin 1	83
PP2S2/RP3S2	Phrasing excerpt 2 / Rhythm excerpt 3 Solo violin 2	93
PP2S3/RP3S3	Phrasing excerpt 2 / Rhythm excerpt 3 Solo viola	78
PP2S4/RP3S4	Phrasing excerpt 2 / Rhythm excerpt 3 Solo cello	73
PP2E/RP2E	Phrasing excerpt 2 / Rhythm excerpt 3 Ensemble	81

Appendix D

PUBLICATIONS BY THE AUTHOR

1. Papiotis, P., Marchini, M., Perez-Carrillo, A., & Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. *Frontiers in psychology*, 5.

2. Marchini, M., Ramirez, R., Papiotis, P., & Maestre, E. (2014). The sense of ensemble: a machine learning approach to expressive performance modelling in string quartets. *Journal of New Music Research*, 43(3), 303-317.

3. Papiotis, P., Marchini, M., & Maestre, E. (2013). Multidimensional analysis of interdependence in a string quartet. *In Proceedings of the International Symposium on Performance Science (ISPS2013)*, Vienna, Austria

4. Marchini, M., Papiotis P., & Maestre E. (2013). Investigating the relationship between expressivity and synchronization in ensemble performance: an exploratory study. *In Proceedings of the International Symposium on Performance Science (ISPS2013)*, Vienna, Austria

5. Papiotis, P., Herrera, P., & Marchini, M. (2013). Aural-Based Detection and Assessment of Real Versus Artificially Synchronized String Quartet Performance. *In Proceedings of the 3rd International Conference on Music & Emotion (ICME3)*, Jyväskylä, Finland

6. Marchini, M., Ramirez R., Papiotis P., & Maestre E. (2013). Inducing rules of ensemble music performance: a machine learning approach. *In Proceedings of the 3rd International Conference on Music & Emotion (ICME3)*, Jyväskylä, Finland

7. Mayor, O., Llimona, Q., Marchini, M., Papiotis, P., & Maestre, E. (2013). repoVizz: a framework for remote storage, browsing, annotation, and exchange of multi-modal data. *In Proceedings of the 21st ACM international conference on Multimedia* (pp. 415-416), Barcelona, Spain

8. Papiotis, P., Marchini, M., & Maestre, E. (2012). Computational analysis of solo versus ensemble performance in string quartets: dynamics and intonation. *In Proceedings of the 12th International Conference of Music Perception and Cognition (ICMPC12)*, Thessaloniki, Greece

9. Marchini, M., Papiotis, P., Perez, A., & Maestre, E. (2011). A hair ribbon deflection model for low-intrusiveness measurement of bow force in violin performance. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2011)*, Oslo, Norway

10. Papiotis, P., Marchini, M., Maestre, E., & Perez, A. (2012). Measuring ensemble synchrony through violin performance parameters: a preliminary progress report. *In Intelligent Technologies for Interactive Entertainment* (pp. 267-272). Springer Berlin Heidelberg.

11. Papiotis, P., Maestre, M., Marchini, M., & Perez, A. (2011). Synchronization of intonation adjustments in violin duets; towards an objective evaluation of musical interaction. *In Proceedings of the 14th conference in Digital Audio Effects (DAFx-11)*, Paris, France

12. Papiotis, P., & Purwins, H. (2010). A lyrics-matching QBH system for interactive environments. *In Proceedings of the 7th Sound and Music Computing Conference (SMC2010)*, Barcelona, Spain

13. Papiotis, P., & Papaioannou G. (2010). Kettle: A Real-time model for Orchestral Timpani. *In Proceedings of the 7th Sound and Music Computing Conference (SMC2010)*, Barcelona, Spain

14. Papiotis, P., & Purwins H. (2010). Real-time Accompaniment using lyrics-matching QBH. *In Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)*. 279-280, Malaga, Spain

15. Papiotis, P. (2010). Real-time Accompaniment Using Lyrics-Matching Query-by-Humming (QBH). *Master thesis*, Universitat Pompeu Fabra, Barcelona, Spain