Multi-modal coupling in musical performance

Davi Alves Mota

Submitted in partial fulfillment of the requirements for the degree of Doctor in Music.

Texto apresentado ao Programa de Pós-graduação da Escola de Música da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do diploma de doutor em música.

Linha de Pesquisa: Sonologia

Advisor: Maurício Loureiro
Co-advisor: Rafael Laboissière

Escola de Música
Universidade Federal de Minas Gerais
Belo Horizonte, June 2017

M917m Mota, Davi Alves

Multi-modal coupling in musical performance [manuscrito]. / Davi Alves Mota. – 2017.
101 f., enc.; il.

Orientador: Mauricio Loureiro.

Área de concentração: Sonologia.

Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Música.

Inclui bibliografia.

1. Música - execução. 2. Comunicação não verbal . I. Loureiro, Maurício de. II. Universidade Federal de Minas Gerais. Escola de Música. III. Título.

CDD: 780.071

This work has been partially funded by a doctoral scholarship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Financial support was partially granted by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais). Part of the research presented in this work has been carried out at the LPNC (Laboratoire de Psychologie et NeuroCognition UMR5105, Université Grenoble Alpes), during a research internship, also funded by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) through a PDSE (Programa de Doutorado-sanduíche no Exterior) scholarship.

Abstract

The manipulation of acoustic parameters in musical performances is a strategy widely used by musicians to demonstrate their expressive intentions. In the case of instrumental ensembles, the coordination of these manipulations between musicians is crucial for the realization of the performance. During a group performance interpreters make use of visual and acoustic information transmitted continuously by other interpreters in order to improve their synchronization. We propose to investigate the interpersonal synchronization/coordination in small musical groups and the role of gestural/visual communication among the members of the ensemble. In pursuit of this goal, we adopted two complementary approaches, the analysis of the body movements of the musicians and the analysis of acoustic parameters extracted from the audio signal.

One could argue that, in the case of ensemble performances, visual information have a crucial role in the coordination of musical events. However, players are also able to follow other musicians solely by listening to acoustic signals. This performance condition has become trivial as a consequence of the growth in music industry, where the demand for perfection and the need to lower the costs led to an increase in the number of recordings made in studio. Moreover, different types of body movements not directly linked to the production of sound can be constantly observed in any instrumental performance, even in *solo* performances. In fact, as pointed out by Gabrielsson (2003, pag. 249), the movement performed by a musician, in addition to communicate relevant information to the coordination with others, may also assume many different roles, such as to communicate expressive intentions, to provide information about the artist's personality or simply entertain the audience. Over the past years, great efforts have been devoted to the study of these movements and, although there is no consensus on its origins or roles, its existence is undeniable (WANDERLEY; DEPALLE, 2001).

This thesis is inspired by findings of previous works (KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012), which argue that musicians playing in ensemble try to adapt their actions to the actions of others, and that the quality of this adjustment could be affected by their movements. In the work we try to review different aspects that integrates the construction of ensemble performance and map their relevance for the adjustment between musicians.

Resumo

A manipulação de parâmetros acústicos em performances musicais é uma estratégia amplamente utilizada pelos músicos para demonstrar suas intenções expressivas. No caso de conjuntos instrumentais, a coordenação dessas manipulações entre os músicos é crucial para a realização da performance. Durante uma performance em grupo, os intérpretes fazem uso de informações visuais e acústicas transmitidas continuamente por outros intérpretes para melhorar sua sincronização. Neste trabalho, propomos investigar a sincronização/coordenação interpessoal em pequenos grupos musicais e o papel da comunicação gestual entre os membros do conjunto. Para tanto, adotamos duas abordagens complementares, a análise dos movimentos corporais dos músicos e a análise dos parâmetros acústicos extraídos do sinal de áudio.

Pode-se argumentar que, no caso de performances em grupo, a informação visual tem um papel crucial na coordenação de eventos musicais. No entanto, músicos também são capazes de acompanhar outros músicos escutando apenas sinais acústicos. Esta condição de performance tornou-se trivial como consequência do crescimento da indústria musical, onde a demanda por perfeição e a necessidade de baixar os custos levaram a um aumento no número de gravações feitas em estúdio. Além disso, diferentes tipos de movimentos corporais não diretamente ligados à produção de som podem ser constantemente observados em qualquer performance instrumental, mesmo em performances *solo*. De fato, como apontado por Gabrielsson (2003, pag. 249), o movimento realizado por um músico, além de comunicar informações relevantes para a coordenação com os outros, pode também assumir diferentes papéis, como comunicar intenções expressivas, fornecer informações sobre a personalidade do artista ou simplesmente entreter o público. Ao longo dos últimos anos, grandes esforços foram dedicados ao estudo desses movimentos e, embora não haja consenso sobre suas origens ou papéis, sua existência é inegável (WANDERLEY; DEPALLE, 2001).

Esta tese é inspirada por resultados de trabalhos anteriores (KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012), que argumentam que os músicos ao tocar em conjunto tentam adaptar suas ações às ações dos outros e que a qualidade desse ajuste pode ser afetada também por seus movimentos. Neste trabalho tentamos rever diferentes aspectos que integram a construção de uma performance em conjunto e mapear sua relevância para o ajuste entre músicos.

Table of Contents

List of Figures

List of Tables

1

Introduction

ENSEMBLE music performance is one of the most challenging tasks in the art world. Some of this difficulty can be attributed to the fact that each musician has an individual way of reading the musical text and the resulting need to coordinate the different conceptions of each group member, which usually turns out to be a very complex task. Additionally, the lack of agreement about the directions the performance will follow can be very frustrating and time consuming. Historically, as musical groups became larger the interpretative choices ended up being centralized, initially in the figure of the ensemble leader and, lately (around the beginning of the nineteenth century), on the conductor, who is entirely dedicated to directing the musical performance and did not play any instrument. This centralized leadership allowed more workable rehearsals and faster results, bringing more unity and discipline for the musical ensembles (SPITZER; ZASLAW, 2004). However, this came at the cost of partially restraining musicians' individuality. While analyzing aspects of leadership in the musical realm, Atik (1994) points out that the career of orchestra musicians basically consists of being constantly told what to do by the leaders, and that they often find themselves in the position of suppressing their needs for individual expression in order to promote the collective good. Asking musicians what makes an outstanding musical ensemble often leads to the same standard answer: "... there are countless factors". One of the factors that is often refereed to is the group unity, the sense of togetherness experienced by the group members. It seems that this sense of togetherness is what reverberates the idea that, such as in a sports team, one individual of the musical group cannot not solve all the problems alone.

The idea that the members of a musical ensemble need to have more experience with each other for accomplishing a better, tighter performance is largely propagated, even though the process of building a performance can go much further than purely musical issues. For example, one may think about the complex cultural structures involved in the formation of independent small chamber groups, and how this can influence from the choice of their repertoire to the choice of the theaters they should perform. There is also the issue of the individual expertise of each performer. A musician can be exceptional, but if she/he is not willing to aid the others, or does not have the experience performing while listening and responding to the musical ideas of others in real time, the homogeneity of the group can be jeopardized. In this context, musicians often describe some sort of musical empathy, that makes possible interprets who never met before to perform together in a coordinated, balanced and natural way. On the other hand, musicians also describe the opposite situation where, no matter the expertise of the colleagues, a cohesive performance is impossible to achieve. In some cases it is easier to give up your own musical identity, sometimes in favor of just one the members and not the entire group, as a way of moving forward with the performance. In short, the most fundamental requirement for performing in ensembles is that the individual parts fit together (GOODMAN, 2002). That is, in musical ensembles the coordination of co-performers actions is paramount for the realization of the musical task.

1.1 INDIVIDUALITY AND ENSEMBLE PERFORMANCE

Studies in music performance suggests that musicians actively manipulate acoustical and temporal parameters in order to express their interpretation of the musical text (GABRIELSSON, 2003). To accomplish this musical modeling, performers manipulate variables within the acoustic limits of their instrument (PALMER, 1997). This subject has been a matter of discussion on several works in the latest years, although not much is known about the mechanisms involved in its coding and decoding (AMELYNCK et al., 2014). When performed consistently, these manipulations can be recognized as a style or “signature” of the interpreter. Often, this musical signature alone is enough to allow the identification of musicians (REPP, 1992). However, musicians playing in ensembles need to coordinate their musical actions with others to achieve cohesion. To do so, performers tend to adjust musical parameters jointly, in search for sound homogeneity (GOODMAN, 2002). The quality of this adjustment appears to be dependent on numerous factors, but it seems to be somehow connected to the way individual manipulations are performed.

Since the studies conducted by Seashore (1938), evidence of remarkable consistency on expressive deviations have been corroborated by a number of other works. Later studies have also observed similarities of expressive deviations in performances of the same piece played by different musicians. In his seminal study on performances of Schumann's Träumerei, Repp (1992) was able to discriminate "commonalities" among interpretations of well-known pianists while evidencing their "individualities". Some of them of irrefutable quality such as the legendary interpretations of Horowitz and Alfred Cortot. Also, according to Repp (1997) individual musicians will differ in the extent they deviate from conventional norms of expression of a particular style, as well as on which resources they use and whether they do so consciously or involuntarily.

As in *solo* performance, musicians playing in ensembles also communicate their expressive intentions through variations in musical parameters, with the additional challenging task of coordinating their actions with co-performers. This is essential for converging to musical cohesion, in which not only note synchronization is achieved, but also musical ideas are coordinated. In ensemble performance, variations have to be matched across partners and individual variations give place to covariation of musical parameters. Other than abilities for anticipating and responding to expressive manipulations carried out by ensemble partners, artful, expressive ensemble performance requires expertise and musical flexibility for producing deviations that allow the statement of her or his individuality but still adhered to accepted norms of expression. Therefore, the greatest coordination challenge in ensemble performance might be to balance one's own interpretation plans with those of the others, either when a leadership role is assumed, i.e. acting as a reference for other players, or following the musical intentions of others.

By manipulating the acoustic parameters of a musical performance musicians are able to manifest their expressive intentions. In instrumental ensemble practice, the control of acoustic parameters is crucial for both acoustic realization of the desired interpretation and the proper understanding of the musical ideas by the listeners. In a musical ensemble, musicians share the responsibility of shaping the sounds, either by serving as a reference for other players (e.g. a conductor) or by following the cues indicated by the ensemble leader. Some studies indicate that musicians anticipate their manipulation of the acoustic parameters by taking advantage of both visual and acoustic information, which is continuously transmitted by other interpreters in order to improve their synchronization and overall musical expressiveness coordination. This study attempts to understand how this fine coordination is deployed by musicians during ensemble performances, in both acoustical and gestural domains.

In this work we try to introduce the concept of "gestural signature", in consonance with the concept of "musical signature", which is widely used in the literature to describe an individual set of characteristics responsible for defining a style of a composer or a performer. The main contributions of this study, therefore, are (1) to demonstrate the existence of those signatures in real life performances and show how they are linked to the so-called music signatures; (2) how those signatures evolve during a duet performance, where the musician in the role of the follower will have to give away his own signature for that of the leader. The concept of signature used in this text describes a group of characteristic marks that serves to set apart or identify a individual, in our case the performer musician.

1.2 DEFINITION OF THE PROBLEM

To introduce the context of this work we may envisage, as a metaphor, the realization of a computerized musical accompaniment system. At the present, no system like that is capable of interacting with other musicians whilst providing the accuracy and the "musical feeling" of a human player. One of the reasons for that lays on the lack of understanding of human-to-human musical interaction. Ideally, to afford a system like that we should first be able to model human-to-human musical interaction by grasping the whole set of, surely multi-modal, dimensions that are enrolled in the process of creating collaborative musical performances. This proposition has the potential of yielding a number of studies in several scientific fields, ranging from musical performance to computer science passing through cognition and psychology, and of course social/cultural contexts.

1.3 HYPOTHESES

Previous studies have shown that musicians actively try to adjust the acoustic parameters of their performances. Other studies presented evidence of coupling between sound and the body movement of listeners and musicians. For example, experiments conducted by Caramiaux, Bevilacqua and Schnell (2010) demonstrated that parameters extracted from the gestures of listeners (position, velocity, normal accelerations) were correlated with acoustic parameters (Loudness and Sharpness) extracted from the sounds they heard. In another study conducted by Dahl and Friberg (2007) participants watched and rated silent videos of percussionists playing under different emotional intentions (happy, sad, angry, and fearful). The results demonstrated that the gestures of percussionists influenced the perception of expressive intentions by the audience.

Given these and other results, it would be possible to observe some kind of relationship between performers' body movement and the acoustical result produced during group performances. Therefore, the hypothesis supported by this study is that, when playing in ensemble, musicians not only try to adjust acoustic parameters, but also tend to synchronize/coordinate their body movements with those of others. This would happen even for movements that are not essential for sound production (accompanying gestures), since they would carry information about the interpretative intentions of the performers, therefore, serving as a means of communication. The formulation of our hypothesis, which is **accompanying gestures carry information about the interpretative intentions of the performers**, is based on observations made in previous studies (KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012), which suggested that musicians try to adjust the acoustic parameters of their interpretation in order to adapt their musical actions to the actions of co-performers; and that the quality of this adjustment is influenced by their movement (GOEBL; PALMER, 2009).

1.4 OBJECTIVES

The aim of this study is to investigate the synchronization/coordination process in small musical ensembles through the acoustic and gestural responses collected during real musical performances. In this work we try to model musicians interactions in typical performance scenarios, using visual and acoustical responses collected during a series of recording sessions. As a consequence, we expect to map acoustic and gesture-related elements that could influence the adjustment between musicians. Regarding the inherent multi-modal complexity in music ensemble interactions, we propose to divide this research into different investigations, which used complementary approaches involving the analysis of kinematic and acoustic parameters. Each investigation is related to a publication realized during the course of this research and is focused on one of the following topics:

1. Modeling intra- and inter-performer synchronization in clarinet duos
2. The effect of leader consistency in ensemble cohesion
3. Gestural signatures formation and disturbance during ensemble performances
4. The adjustment of acoustic parameters and its reflection on musicians gestures

1.5 RELEVANCE

Currently, we have witnessed a rapid growth of interest in the relationship between body movement and human communication. This growth was driven especially by the development of new technologies of motion capture, which significantly expanded the perspectives of researchers in this field. In the context of music research, several studies have demonstrated the relevance of human movement as a means of communicating intentions and emotions in musical performance. Despite great attention paid to the issue, the processes behind the relationship between gesture and music are still not clear. By investigating the multi-modal relations occurring in musical performance we hope we can help to clarify the process of construction of a musical interpretation and demonstrate how gesture information contributes to it. Furthermore, the recurrence observed in the musicians' gestures strongly suggests that encoded musical information can be somehow embedded on them. However, the interaction between gesture and music points to a complex relationship, dependent on innumerable variables. A closer investigation of this relationship may reveal underlying mechanisms involved in collaborative musical interpretation.

1.5.1 *Implications of results and/or applications in musical practice*

Applications of this research may emerge for educational environments where computational visualization tools can help raise awareness of the movements being performed and their effect on the coupling of the sound between musicians. Also, new interfaces for digital musical instruments, or intelligent accompanying systems, able to "follow" gestural and acoustic cues using communication strategies similar to those employed by real musicians.

1.6 OUTLINE OF THE RESEARCH

As stated before, this work was accomplished by conducting different complementary studies. By using gestural and acoustical information, each of these studies aimed at different aspects of musician's interactions in typical performance scenarios, in an attempt to build a model for ensemble performance interaction. First, **Chapters 1** and **2** try to define the focus of the study by presenting its objectives and hypotheses, followed by a review and discussion of the relevant literature for the work. **Chapter 3** presents an overview of the experimental procedures and musical materials used in the work, including a systematic review of modern note identification methods and a discussion about the validity of those methods for synchronization studies. The following chapters present the studies conducted during this work, which are grouped in two main approaches:

- **Studies addressing the acoustical component of musical ensemble performance (Chapter 4)**

  Section **4.1** discuss the synchronization between musicians and how different leaders may influence in the adjustment of the duo, after shortly reviewing the theoretical basis of the self-other effect in synchronization of musical tasks.

  Section 4.2 discuss the concept of rhythmic signature, its formation and the effect of the leader rhythmic consistency in the synchronization of the duo.

- **Studies addressing the gestural component of ensemble musical performance (Chapter 5)**

Section 5.1 discuss the concept of gestural signature, how it can be use to identify musicians and, moreover, how musicians tend to keep their original gestural signature while following their own recordings.

Section 5.2, focus on how musicians change their gestural signature while following other musicians and, furthermore, tend to bend their gestures patterns towards the leaders'.

- Lastly, **Chapter 6** summarizes the achievements of the research and discuss the combined implications of the results and future studies derived from this work.

2

Background of the research

In this chapter we present a brief review of studies with historical relevance for the synchronization of musical performances and draw an overview of the theoretical background supporting the hypotheses proposed in Section 1.3. We focus on studies regarding the investigation of cognitive processes involved in the collective musical practice. Further, we present an overview of the research in musical gesture, outline the concepts of gesture and gestural signature used in this work and briefly discuss how gesture-music interactions can influence in the coordination of musical ensembles.

2.1 OVERVIEW OF ENSEMBLE SYNCHRONIZATION RESEARCH

One of the earliest studies investigating synchronization in musical ensembles is Rasch (1979). It presents a framework of time measures for describing what the author calls "asynchronization" in ensemble performances. In this work, synchronization was measured with specially designed recording and analysis methods, using directional microphones in an anechoic room. The narrow angle of sensitivity of the microphones and the lack of reverberance in the room allowed a good separation of the sound sources which, in turn, were feed to a analog-to-digital converter and to a computer where the envelop of the signals were calculated. The onsets were estimated as the point where the signal reached a threshold of 15 dB below the maxima of each note. The collected data showed that asynchrony, defined as the standard deviation of differences in onset time of simultaneous notes, had typical values of 30 to 50 ms. The work also discuss the implications of these findings for perception and performance, citing one of his previous studies about the perception of polyphony (see: Rasch (1978)) to discuss how the chosen energy threshold and other factors such as the perception of tone order and motor control aspects of the bowing technique of violin players could influence in the results. One of the key problems for synchronization studies is the estimation of the onset instants. The energy-based onset detection procedure applied in this study is fairly reliable, although the method disregards the influence of attack times in the perception of note onsets, thus, lacking the precision of modern methods. Nowadays, most note segmentation systems make use of multiple spectral features extracted from the sound signal in order to estimate the segmentation points which, in turn, provide further robustness to the procedure.

Shaffer (1984) describes studies of timing in solo and duet piano performances, in which musicians gave repeated performances of the music. In both solo and duet performances there was expressive use of timing, modulating the tempo of the music and the phase relationship between the voices. The expressive forms were similar in successive performances of the piece. (HURON, 1993) perform an analysis of 15 two-part inventions by J. S. Bach to show that the amount of onset synchrony is significantly less than would be expected by chance, and so suggests that synchronous onsets are intentionally minimized by the composer. The results are consistent with the objective of maintaining the perceptual independence of the polyphonic voices. Keller (2001) presents the theory of Attentional Resource Allocation in Musical Ensemble Performance (ARAMEP), which accounts for how attentional flexibility is influenced by various musical and extra-musical factors, a cognitive model of attention allocation for ensemble performances. Goodman (2002) reviews aspects of ensemble performance: coordination, communication, the role of the individual and social factors. Glowinski et al. (2014) investigated expressive non-verbal interaction of musicians in a string quartet, performing a piece of Schubert in two different conditions. The first, in a concert-like situation and the second in a perturbed situation, where the first violinist gave alternative interpretations for the score without the knowledge of the others performers. Results show that musicians' head movements were higher in the perturbed condition, and that musicians pay more attention to other performers' heads to better predict their upcoming actions.

Marchini et al. (2014) describes a novel method for building computational models of ensemble performance. Using machine learning algorithms to produce models for predicting musical parameters manipulations. Bishop and Goebl (2014) investigated the potential instrument specific effects of expertise on sensitivity to audiovisual asynchrony among highly-skilled clarinetists, pianists, and violinists, using mismatched video clips of performances. In their pilot experiment participants were asked to view video clips and indicate as quickly as possible whether the audio and video were from the same performance or different performances by pressing one of two marked keys on the computer keyboard. Results suggest that participants detected audiovisual mismatches most readily in violin stimuli and least readily in piano stimuli. In their main experiment participants were asked to view each item and indicate as quickly as possible whether or not the audio and video were synchronized by pressing one of two marked buttons on the computer keyboard. Results indicated that perceptual-motor expertise relates to improved prediction of observed actions.

2.1.1 *Skilled joint action performance and Perception-action Coupling*

As previously mentioned, is common sense that in musical ensembles the coordination of co-performers actions is paramount for the realization of the musical task. A central question for this work is how do we manage to become effective in performing such coordinated actions, either when we play as leaders or as followers. The problem of action coordination is not exclusive to the musical field, it can be observed in different types of human interactions carried out in tasks such as in groups of people dancing, in the practice of collective sports or even ordinary everyday situations such as washing dishes with a partner. In the literature this process is often refereed to as Skilled Joint Action Performance which is defined by Sebanz, Bekkering and Knoblich (2006) as a moment "[...] when two or more people coordinate their actions in space and time to bring about a change in the environment". But how do people coordinate their actions? A popular hypothesis from cognitive psychology supports that the act of anticipating the effects of other peoples actions may be in the center of this process (KNOBLICH; FLACH, 2001). In the light of this hypothesis, the processes of perception and action would be connected through the underlying cognitive functions common to both systems (JEANNEROD, 2001; WOLFGANG, 1997; SEBANZ; BEKKERING; KNOBLICH, 2006). This perception-action coupling process implies that observing the act performed by another person activates a simulation corresponding to that action in our own motor system (LEMAN; NAVEDA, 2010; KELLER; KNOBLICH; REPP, 2007) and, accordingly, would be strengthened as a result of the training process (NOVEMBRE; KELLER, 2014).

This is the fundamental assumption of the so called *Common-coding Theory* (HOMMEL et al., 2001), which holds that actions are coded in terms of the perceptible effects they should generate. According to this theory, the observation of an act performed by another person would activate a mental simulation corresponding to that action in the motor system of the observer. Thus, similar motor representations would be activated in the observation and the production of actions. Some studies in neurophysiology pointed to evidence that this coupling between perception and action could be implemented at a neuronal level. For instance, an experiment performed by Gallese et al. (1996), where neuronal activity of monkeys was recorded during the observation and the realization of simple tasks, indicated that the pre-motor cortex neurons of the animals were activated not only when the monkeys realized movements to grab objects, but also while observing the experimenter performing the same acts. A more recent study showed evidence of the same process taking place in a musical task. Meister et al. (2004) asked piano students to participate in functional magnetic resonance imaging (fMRI) sessions in two conditions, playing an excerpt of a Bartok piece with their right hand on a plastic soundless keyboard and imagining themselves playing the same piano piece. The results indicated that the neuronal activations during the imagery condition corresponded to those in the music performance condition.

2.1.2 *Distinction between first and third person, the* self-other *effect*

One consequence of the interaction between action and perception cognitive functions is that it can contribute to the prediction of future outcomes of currently perceived actions. Furthermore, if the internal models formed in previously performed actions are applicable to currently observed actions, we should be better in predicting the results of our own recorded actions than the results of actions performed by others (KNOBLICH; FLACH, 2001). In Knoblich and Flach (2001) the authors asked the subjects to watch videos of themselves and other trowing darts at a target board and, later, to predict the darts landing position. The results indicated that the participants were more accurate in predicting the outcome of their own actions. This hypothesis implicates that the previous experiences could be of crucial importance to the results obtained in a collaborative task. For instance, while learning how to pass a ball to a colleague during a soccer match one may be subjected to different informations that should be relevant for the internalization of the cognitive model, such as the weigh of the ball, the roughness of the ground, the air resistance and her/his own kicking strength. When the situation is inverted and the person has to receive the same pass this model scheme would serve as a reference frame for all the internal calculations that would enable this person to accurately position her/himself to receive the pass. Hence, we can imagine that the more variables in the processes stay unchanged, the more predicable the outcome would be, and this would include information related to the other person's body, like their kicking strength. The rationale behind this exemple could be extended to all tasks where the prediction of another person's actions are required, even the musical tasks.

The self-other effect, therefore, flirts with the idea that you yourself may be your best co-performer in terms of achieving a tighter coordination in an ensemble performance. The synchrony between musicians playing together would result from the simulation of the action performed by the other. That is, during the group performance, the musician would simulate how the accompanying instruments (the parts of the others) should be played. This process would be independent of how the musician is playing his part. The simulated parts would evoke the characteristic traits of the musician (articulations, variations of dynamics and agogic, etc.), reflecting how this musician would execute that part. As a result, the musician, when accompanying himself, should be able to predict more accurately any significant variations introduced in the performance. In addition, this hypothesis implies a possible correlation between synchronization accuracy and self-recognition, that is, musicians that recognize themselves more easily would also be able to better synchronize with themselves. Some studies approached this topic by testing how musicians react while following their previously recorded performances. For instance, Keller, Knoblich and Repp (2007) recorded pianists playing both parts of a duet and later asked them to follow recordings made by themselves and others. Results indicated that participants were not only better synchronized with their own recordings, but also were capable of recognizing themselves very accurately. Mota (2012) and Mota, Loureiro and Laboissière (2013) replicated this study using clarinet duos, achieving fairly similar results.

2.2 GESTURE AND MUSIC

Body movements can be observed in any instrumental performance. Some of those movements are essentially linked to sound production, like the movement of the fingers of a pianist, or the wrist of a violinist. However some of those movements cannot be linked to sound production, like body sway or head movements. Gabrielsson (2003, p. 249) suggests that movements performed by musicians may take several roles besides communicating relevant information to the coordination with others. He states that they can also serve to communicate expressive intentions, provide information about the personality of the artist or simply entertain the audience. In recent years, great attention have been devoted to the study of this kind of movements and, although there is no consensus about its origin or function, its existence is undeniable (WANDERLEY; DEPALLE, 2001). Based on the observations made by Delalande (1988) Cadoz, Wanderley et al. (2000) proposed to differentiate body movements that are directly related to the production of sound (instrumental gestures) from those that are not (ancillary gestures). They suggested that the latter present tighter relations to the performer's expressive intentions, making inferences about the role played by each one in musical performances (see also: Wanderley (1999), Jensenius et al. (2010)). To support this theory, the authors analyze the interaction between experienced musicians and their instruments by reviewing the use of the term gesture in the musical and human-machine interaction domains. Their aim is to propose a discussion about the different classifications of gestures, presenting topics from other disciplines that may be relevant to the discussion about gesture and music. This study is one of the first attempts to define a typology of the musical gesture. Other works like Wanderley et al. (2005), Rasamimanana (2012), Desmet et al. (2012) attempted to characterize and quantify physical gestures involved in musical performance, in order to identify their musical significance.

Authors like Eduard Sievers (1850-1932), Gustav Becking (1894-1945) and Alexander Truslit (1889-1971) pioneered the combined analysis of music and gesture. The early interest in the subject departs from the concept of "auditory motion information", in which the movement information would be encoded in the expressive microstructure of the performance (REPP, 1993b, p. 168). The first attempts to empirically investigate the relationship between movement and music dates from the early twentieth century, made by Gustav Becking in 1928, and Alexander Truslit in 1938. These authors have focused their work on the assumption that the music could be described by gestural information. As pointed out by Repp (1993b), their work present results of a series of experiments aimed at the reconstruction of gestural information contained in music. Gustav Becking (1894-1945) devoted himself to the systematic study of the music of several composers, through the characterization of the musical tempo extracted from diagrams drawn by subjects performing gestures, similar to those of a conductor, while listening to musical excerpts. Becking intended to extract meaning from the comparison of curves extracted from similar musical excerpts. Nettheim and Becking (1996) presents a synthesis of the book *Der musikalische Rhythmus als Erkenntnisquelle* written by Becking in 1928, pointing out the similarities between Becking's work and more recent research on gesture. The book of the musicologist Alexander Truslit (1889-1971) *Gestaltung and Bewegung in der Musik* contains reflections on the nature of the musical gesture and its role in musical performance, part of these speculative and lacking scientific rigor. However, the book presents relevant ideas for the most contemporary research on the field, acting as a source of hypotheses for more precise questions, whose approach has become feasible only recently due to the technological advances that have enabled the accurate collection of gesture data. A comprehensive review of Truslit's work can be found on Repp (1993a). The author, Bruno Repp, is the main responsible for the dissemination of the book, which was previously restricted to small groups of researchers.

Clynes (1995), proposed a review of Becking experiments making use of a device invented by himself called *The Sentograph*, used to detect variations in the pressure exerted by the tip of the listener's finger while accompanying the music. As Becking, Clynes focused on describing the nuances found in performances of the works of some composers, proposing what he called *composers' inner pulse* as a way to determine specific curves for each composer studied. Clynes argued that the meaning of music would come from *essentic forms* – dynamic curves that characterize basic emotions, defined by the musical structure, and that the *composers' inner pulse* should somehow manifest in the expressive microstructure of the performance. In the study, data were collected from five groups of subjects with different levels of musical proficiency, including renowned musicians such as Vladimir Ashkenazy and Yehudi Menuhin, music students at various levels of experience and non-musicians, totaling 135 subjects. The subjects listened to forty four-bar excerpts extracted from compositions of Beethoven, Mozart, Haydn and Schubert, executed by a computer, in which the "inner pulse" of the different composers was incorporated almost randomly, so that for every four performances one would incorporate the "correct" pulse and three would incorporate the "wrong" pulse. The subjects were instructed to indicate their preferred performances. The results showed that the greater the musical proficiency the listeners had, the higher the preference for "correct" performances. The results obtained by the author are controversial because of the lack of scientific rigor and internal inconsistencies in the theory of the composer's personal pulse. Nevertheless, the article is relevant in the historical context of research in musical gesture because of its methodological initiatives, which involve concepts from different areas such as experimental psychology and neuroscience.

In the early 1990s, Neil Todd proposed structure-level models of expressive performance, based on observations of rhythms variations (TODD, 1989a; TODD, 1989b) and intensities (TODD; NEIL, 1992). Unlike Becking and Clynes, Neil Todd focused on the mental representations by making analogies between rhythmic/dynamics variations and physical movement. In more recent studies, data extracted from these models were compared with time and dynamic curves produced by interpreter's head movements during piano performances. The results indicated a partial similarity between the actual and model curves, but no quantitative evaluation was performed (REPP, 1993b; WIDMER; GOEBL, 2004). The results led the author to suggest the hypothesis that there would be a direct interaction between the motor system and auditory system, so that internal representations would be evoked directly in the cerebral motor center by sensory stimuli that correspond to these representations. This would indicate a close connection between the body language and musical expression, as the performers hear their own performances (TODD, 1999). Dahl and Friberg (2007) indicate that the relationship between music and movement can be described by different aspects. Of these, the most notable is the fact that the sounds of traditional acoustic instruments are produced from the human motion, so that specific motion characteristics will be inevitably reflected in the resultant sound.

2.2.1 *Gesture in ensemble performance*

It is also well known that body movements can communicate interpretative intentions in music performance. In ensemble performance, ancillary body movements may provide a valuable resource for improving communication of intended expression, which facilitates musical coordination. According to Keller (2014), ancillary movements "[...] generate kinesthetic feedback that aids the performer in regulating technical and expressive parameters of sound production" (p. 268). On a previous study, Keller (2008) suggested that the coupling of such movements in ensemble performance might reveal the quality of interpersonal coordination. The author proposed three ensemble skills required for achieving the remarkable precision and flexibility of ensemble performance coordination. One of them, the ability for anticipatory auditory imagery, assumes that such anticipation involves mixtures of auditory and motor imagery.

In a more recent study, Keller (2014) argues that musicians playing in ensemble activate anticipatory musical imagery, "[...] an advanced cognitive-motor skill that is refined through musical experience" (p. 273), which functions as internal representations for their own interpretation plans, as well for the predictions of those of the partners. He claims that abilities to use anticipatory imagery to predict the actions of co-performers facilitate interpretation, planning and execution and may determine the quality of ensemble cohesion. In a study on piano duets conducted with expert pianists, Keller, Knoblich and Repp (2007) observed that participants predicted better the results of their own recordings, hence the author's suggestion that upcoming events should have been simulated by the cognitive-motor system.

In contrast, only a few studies have addressed the relation between expressive coupling and body movement in ensemble performance. Aiming at investigating the cognitive-motor skills that mediate interpersonal interaction in musical ensembles Keller and Appel (2010) were able to identify systematic relations between sound synchrony and interpersonal body sway in piano duos. Goebl and Palmer (2009) investigated timing and synchronization aspects as well as finger kinematics and head motion in duet piano performance. They showed evidence of systematic interactions of both timing and motion between co-performers. They also observed an increase in musicians' body movements when auditory feedback was reduced, particularly, finger heights above the keys and head movements, which became more synchronized, even with note synchrony decrease.

2.2.2 *Gestural signature*

Recurrence of gestural patterns has been observed even in different musical contexts. Wanderley et al. (2005) pointed out that individual performers tended to maintain consistent patterns of movement throughout "standard" and "exaggerated" performances of the same piece. This recurrence was also observed in listeners' free body movement responses to music (AMELYNCK et al., 2014). As stated before, recurrence in the manipulation of musical parameters was observed and fairly discussed in different studies, which supports the concept of a musical signature of performers. In a similar manner, it could be possible that musicians would tend to exhibit recurrence on body movement patterns as a gestural signature. The occurrence of gesture signatures in everyday tasks has been the objective of several studies, some examples are: Farella et al. (2006) and Loula et al. (2005). Likewise, recurrence in gestural patterns was also observed in musical performances, where musicians consistently reproduce the same pattern of gestures while executing similar musical content (WANDERLEY et al., 2005; NUSSECK; WANDERLEY, 2009).

2.3 GESTURE-MUSIC AND ENSEMBLE COORDINATION

As previously stated, in ensemble music performance synchronization can be achieved by means of acoustical or visual streams, but a number of studies shows that the combination of these sensory modalities can improve the overall synchronization of any ensemble. A study conducted by Nusseck and Wanderley (2009) indicated that changing the kinematic properties (amplitude of the movements) of a musician movements can influence the perceptual impressions of his performance. In the study, the authors used kinematic displays, a stick figure representation of musicians' body. They ask subjects to rate specific music-related dimensions of the performances (Perceived tension, intensity, fluency and professionalism). Results shows that participants judged as more intense the interpretations with the range of motion digitally enhanced, even without any changes in the audio streams. We believe that the coherence between these modalities should play a key role in this process.

Leman (2007) focus on the theoretical coupling between the musical experience (mind) and the audio signal (matter). The author proposes a theory (*embodied cognition theory*) capable of addressing the problem of mediation existing between these two domains, assuming the human body as a mediator biologically shaped to transfer physical energy to a mental level, also acting in the reverse process, transferring a mental representation into the material form. The author argues that, in certain circumstances, the natural mediator (human body) can be extended with artificial mediators, as in the case of a musical instrument or a human-machine interface. The author describes ways to analyze these cases and propose possible applications of his theory, for example, the integration with musical instruments and the retrieval of musical information. The book is a reference in the field of studies of music and movement. It condenses the results of two decades of research on the topic, culminating in the elaboration of a theory that forms the basis of much of the research on musical gesture currently performed.

Dahl and Friberg (2007) suggested that the gestures performed by musicians during a performance would act as a vehicle for the expression of their musical intentions, in order to provide a channel of direct communication with the listener, independently of the auditory information. In support of this hypothesis the authors performed an experiment to explore the transmission of emotional intentions through the gesture of the musician. Subjects were instructed to rate musical performances according to their perceived intentions. To that end, they watched silent videos of musicians playing under four distinct emotional intentions. The videos were presented under different viewing conditions, showing only certain parts of the instrumentalist's body. The results demonstrated that certain emotions were better communicated than others, and that the identification of the intention transmitted was little influenced by the conditions of visualization.

Goebl and Palmer (2009) observed an increase in musicians' body movements when auditory feedback was reduced, particularly, finger heights above the keys and head movements, which became more synchronized, even though note synchrony decreased. Pianists were instructed to assume musical roles as leader and follower. Analyses of the timing variability suggested that playing with another performer affected the follower more than the leader. The follower's timing was less precise when playing with the leader than when playing alone. The follower adapted his or her timing more than did the leader. Keller and Appel (2010) investigated the role of anticipatory auditory imagery in musical ensemble performance, by identifying systematic relations between sound synchrony and interpersonal body sway in piano duos. Their results shows that ensemble coordination was not markedly affected by whether pianists were in visual contact during the experiment. The only hint of such an effect was seen in the higher variability of asynchronies observed when visual contact was present than when it was absent.

Caramiaux, Bevilacqua and Schnell (2010) explores relationships between gesture and sound, by reading the gestural responses produced by subjects while listening to different sounds. To do so, the authors use Canonical Correlation Analysis (CCA) to measure the linear relationship between two sets of variables, extracted from subjects' body movements and from audio samples used in the experiments. Data were collected in experiments where subjects were instructed to perform free body movements while listening to recorded sounds, imagining that the sounds were produced by themselves. The canonical correlation analysis (CCA) clearly demonstrates the existence of some relationship between the variables: the gesture of the subjects and the recorded audio samples. Tsay (2014) demonstrates that across six studies, visual information dominated rapid judgments of group performance. It suggests that coordination in musical ensembles involves the activation of multiple sensory modalities that are assumed to work in a coherent manner. Likewise, it supports that if we assume the existence of musical information encoded on the gesture of a musician, and that this information contributes to the coordination of the ensemble, we can also assume that changes in the kinematic representation of a musician can affect the overall synchronization of their peers musicians.

3
Methods

STUDIES presented in this work were realized with data collected over several recording sessions made in the CEGeME laboratory (Escola de Música da Universidade Federal de Minas Gerais, Belo Horizonte, Brasil). In this section we make an overall description of the recordings sessions and the preprocessing methods applied. Some methods that are specific to certain investigations will be discussed in detail as appropriate. The execution of this work required the development of a multi-modal analysis framework capable of accessing the relationship between the movement of musicians and the acoustical result produced in ensemble performances. Towards the development of this framework, we made use of methodologies that already revealed important aspects of this connection, which where developed in previous studies such as Loureiro et al. (2009), Loureiro, Campolina and Mota (2009), Mota (2012).

3.1 DATABASE COLLECTION

The realization of this work required the acquisition and organization of a performance database that included gesture and sound data aligned to the performed scores. This database had to be recorded simulating typical performance scenarios which were familiar for professional symphony orchestra musicians. The objective was to interfere as little as possible in the typical routine of the musicians during the recordings, by providing a more conventional environment for the data collection, in the hope that this could provide more consistent results with their musical practice.

The choice of instruments was initially guided by the proposal of analyzing body movements that are not essential for sound production. In wind instruments, body movements directly related to sound production, the so-called *effective gestures* (CADOZ; WANDERLEY et al., 2000), are limited to movements of small amplitude, such as fingers, jaw, lips, tongue and chest. This characteristic facilitates and induces the musician to engage in a wide variety of movements of legs, arms, head and postural variations, which are less dependent to sound production and, hence, more related to musical intentions. Moreover, focusing on movements not directly necessary to sound production facilitates movement data aquisition which can be easily done without hampering the playing technique of instrumentalists. An instrument with the aforementioned characteristics allows the collection of movement data without major effects for the instrumentalists, which would otherwise interfere in the results obtained. Furthermore, we have accumulated considerable experience in the analysis of acoustic data produced by woodwind instruments, as demonstrated by Loureiro et al. (2009), Loureiro, Campolina and Mota (2009), Loureiro et al. (2012) and others.

The experimental protocol was designed to simulate the recording of a two-part musical excerpt, where musicians play over the recording of another musician. This performance condition has become trivial for most musicians as a consequence of the growth in the music industry over the last century, where the demand for perfection and the need to lower the costs led to an increase in the number of recordings made in studio. We recruited six musicians, five professional clarinetists and one student to participate in the experiments. Previous experience in orchestra and/or other ensembles was mandatory. They were asked to play two musical excerpts, the first one from Symphony No. 5 in E minor, Op. 64 by Pyotr Ilyich Tchaikovsky, the opening 20 bars of the first movement (Figure 3.1); and the second one from the "Dance of the Peasant and the Bear" from the ballet *Petrushka* by Igor Stravinsky, extracted from the *Quatrième tableau* N° 100, first three bars (Figure 3.2). In both passages first and second clarinets play a solo melody in unison (*soli a 2*), which requires the synchronization of each note. Recordings were performed in two sessions separated by an interval of two days. In the first session, musicians were instructed to play four times as the leader of the duo.

At the end of the session, participants were asked to select their preferred performance to be used in the next session. In the second session participants were asked to play as second clarinetist, following leaders' recordings selected in the first session, including those performed by themselves. They were instructed to follow the leader, as they would do in an actual orchestra rehearsal. Metronome beats were included at the beginning of leader recordings in order to facilitate synchronization of the first note. Tempo was estimated by the detected duration of the notes on the first bar. Participants were recorded in a single run, playing as second clarinetist while listening to leaders' recordings through a headset in one of their ears. They were allowed to listen once to the entire leader execution before recording. The use of the headset was required to prevent the audio from the first clarinet to be recorded along the audio of the second clarinet. Leader recordings used in the second session were presented in randomized order to the participant. Musicians were not made aware of which clarinetist they were following, even when following recordings made by themselves. No visual information was provided.

Figure 3.1: Excerpt from Tchaikovsky's fifth symphony, *opus* 64 (first 20 bars).

Figure 3.2: Excerpt from the "Dance of the Peasant and the Bear" from the ballet Petrushka by Igor Stravinsky, extracted from the *Quatrième tableau* Nº 100 (first three bars).

3.1.1 *Audio data*

Note Identification

One of the key points for studying the synchronization in musical ensembles is the ability to detect note onsets with accuracy. To the present, analysis of timing and synchronization in musical performance studies has been limited by the precision of note detection methods. Although several solutions were proposed along the years, there is no consensus about which method is more appropriate for timing related tasks. To mitigate this issue, in recent years, the majority of works dealing with synchronization are tied to the use of MIDI interface devices (e.g. electronic keyboards) leaving aside acoustic instruments and voice. This poses a problem and leads to a narrow interpretation of what synchronicity means, by completely disregarding its (clearly) multidimensional nature. The concept of musical note itself represents a major issue in moving away from this keyboard centered setup.

There is still a lot of debate over the exact moment in which the onset of a note is perceived. Depending on the method used to segment the signal the resulting asynchronies can vary largely. Energy based methods tends to work better for percussion-like signals but are not well suited for pitched signals. The lack of energy discontinuity in note transitions is actually one of the goals pursued by highly skilled musicians, a feature that makes the use of energy based methods even more difficult. On the other hand, pitch based methods are problematic when dealing with repeated notes. A combination of those methods should be ideal, but the shortage of understanding of how human internal segmentation process work can still lead to inconsistent results.

Although several methods exist for performing the task, there is no consensus about the best or the more precise of them all. A great review about the subject is Bello et al. (2005) and, even though the study is more than ten years old, not much has changed since then. Bello et al. (2005) states that the distinction between the concepts of *attacks*, *onsets* and *transients* is key for the categorization of the onset and, furthermore, this concept may vary depending on the aimed application. The authors describe the *attack* as the region where the energy envelop increases, the *transient* as a short interval where the signal show a non-deterministic behavior and the *onset* as a single instant, chosen by the detection system within the time interval of the *transient* region. Yet, most of the time the chosen onset instant will coincide with the start of the *transients* region. In this definition, the concepts of *attack* and *transients* are often superposed and their disambiguation will depend on the onset detection criteria, which will be responsible for selecting where in the transient region the onset instant will be marked. This is true if the onset method used is aware of the importance of the transients, but most onset detection techniques are energy based methods which demonstrate great robustness for percussive sounds but lack of precision on notes with softer, longer attack times, for instance, woodwinds instruments.

Recently, authors have questioned those limitations and their impact on the obtained results. For example, Novembre and Keller (2014) argues that repetitive movements, as finger-tapping tasks, do not capture a crucial component of human action: its structure. This is causing an expansion of the methods towards more musical-friendly experimental setups. This notion vouches for the use of acoustic instruments in synchronization experiments, but the question of which segmentation methods should be used is still controversial. Some studies tried to apply this approach to the problem, reaching feasible solutions. For instance, Palmer et al. (2013) defined the note onsets as the point after each minimum at which the signal exceeded the minimum by more than 5% of the maximum-minimum difference. In a similar approach, Wing et al. (2014) and Timmers et al. (2014) detected the local maximum of the signal corresponding to successive notes, and note onset times were determined using an adaptive threshold applied to the "valley" preceding each maximum. This event detection method was visually cross-validated with their spectral analysis for the entire data set.

Perceptual onset vs acoustical onset

Another factor to be considered is the distinction between the concepts of perceptual onset and physical onset, i.e. two notes that are physically synchronous may not be perceptually synchronous. In Vos and Rasch (1982) the authors suggested discriminating perceptual and physical onsets by its energy levels. The experimental results demonstrated that perceptual onsets usually takes place in about 6 to 15 dB below the maximum energy value of the note. Nevertheless, Dixon (2006) points out that the work do not take into account factors like masking or temporal order thresholds that usually take place in complex musical works and can definitely influence the onset perception.

A perceptual onset is not always comparable to its physical counterpart, for instance, it is reasonable for the physical onset to be extracted from the audio signal, relying solely on a set of predefined rules. However the estimation of perceptual onsets should consider the influence of subjective factors that could be related to one's individual musical experience. For example Leveau and Daudet (2004), whilst describing a methodology for evaluation of onset detection algorithms by cross-validation with manually annotated databases, points out several elements that can disturb one's decision in manually estimating onset instants on real world performances. Examples are: room acoustics effects (by increasing the release time), polyphony (broken chords can be considered as a note sequence or a block), super-positions of the previous and current notes (e.g. on bowed instruments).

Description of the current method

For this study the audio was captured in one channel using an omnidirectional microphone placed at an approximate distance of 1 meter from the instrument, in a room with basic acoustic treatment. Each clarinetist used his own instrument and materials during recording sessions. Audio tracks were semi-automatically segmented at note level using the tool Expan (for a complete description, see Loureiro, Campolina and Mota (2009)). This tool is one of the projects of the CEGeME laboratory and it is in active development since 2008. It is a live project where developers and contributors are constantly in search for new methods to increase the tool's efficiency. The system uses a combination of spectral and temporal parameters of the audio signal to perform the segmentation: detecting abrupt changes in windowed RMS (Root Mean Square) signal and pitch variations. As said before, the detection of abrupt variations in the amplitude envelope is the most common method used for note onset identification (BELLO et al., 2005). Although this approach is very suitable for percussive sounds, such as drums, the piano or plucked strings instruments, it is very inefficient for instruments with a controllable energy envelope, such as wind instruments (clarinet, oboe, trumpet) or of bowed strings (violin, cello, double bass).

As a way around this limitation our system looks for variations of pitch values greater than ca. 6%, which is approximately the size of a semitone, in order to estimate the onsets. For repeated notes, i.e subsequent notes of the same pith, our system relies on the small energy gap between repeated notes, which causes a more evident variation on the RMS values and, hence, the onset detection. If this variation is not high enough the system will discard the onset. The Expan system performs note onset calculations on windowed signals, defaulting to windows of

1024 samples (23.2 ms in a 44100 sampling rate) with a superposition between windows of 256 samples (5.8 ms in a 44100 sampling rate), which gives the system a precision of ca. 6 milliseconds for onset detections.

3.1.2 *Gesture data*

The system *Optotrak Certus*[1] was used to track the tridimesssional position of two rigid bodies[2], the clarinet bell and the head of musicians. To accomplish this, a group of three markers were used in each rigid body. The sampling rate used for gestural data collection was 100Hz. The first group of markers was fixed on the bells of the clarinets and the second on the heads of the participants, in order to allow the tracking of the six degrees of freedom of the bodies with minimal interference in the movement of the musician - for more details and examples of configurations of similar experiments, see Wanderley et al. (2005), Goebl and Palmer (2009), Keller and Appel (2010). The positioning of markers is demonstrated in Figure 3.3.

Previous studies have shown that the movement of the clarinet bell is responsible for much of the movement performed during the performance of the instrumentalist (WANDERLEY, 2002; PALMER et al., 2009a). For instance, in Wanderley et al. (2005) the body movements of clarinetists and the clarinet were quantified by calculating the differences between successive frames of video recordings. The results showed that 30% of the total movement was performed by the bell, compared to 20% for head movements and 10% for leg movements. Based on this results, we chose to limit our analysis to the movement of the clarinet bell.

The trajectory of the clarinet bell is coupled to the translational movements of the musician's body. This means that any movement produced by up and down, side to side, anterior/posterior and rotational move-

[1]Optotrak Certus (manufactured by North Digital Inc. [NDI]) is a three-dimensional motion capture system that uses infrared light emitting diodes (LEDs) as active markers. These markers are individually identified by a frequency of optical pulses and their positions are measured by a trinocular camera system. The system can simultaneously measure up to 512 individual markers or a large number of rigid bodies with markers embedded.

[2]Rigid body is a conceptual representation of a solid body of finite size in which the deformation is disregarded. That is, the distance between any two points of a rigid body remains constant over time regardless of the external forces exerted on it. The spatial position of a rigid body can be represented by the coordinates of a body reference point (usually coincident with the centroid of the solid), and its angular position, described by the rotation about the reference axes.

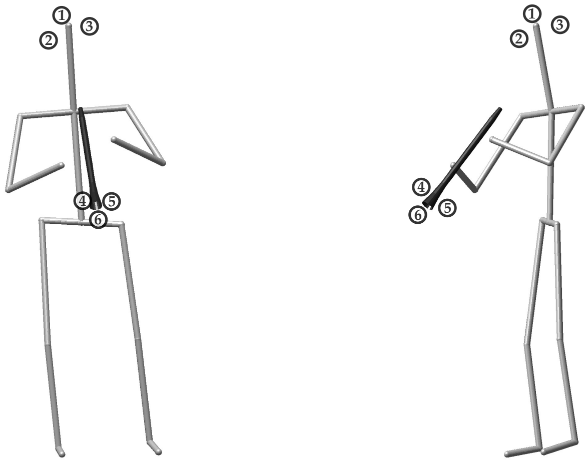

Figure 3.3: Kinematic representation of a clarinetist body, showing marker positions on the head and the clarinet bell.

ments can be reflected in the clarinet trajectories. This could be interesting if we wanted to describe the clarinet bell's movement from the perspective of the audience or the co-performers. Although this combination of body and bell movements could contain relevant information for the synchronization between musicians, it can be very problematic to find which body part is influencing the bell's trajectory. For example, some observed bell movements could be caused both by knee flexion, arm movement or any combination of them. Moreover, it can also increase the amount of movement in bell's trajectories and variations between takes making it more difficult to identify possible gestural recurrences. A solution to work around this problem would include collecting the movement of different body parts in order to find and link the source of movement to a possible musical counterpart. Although examples demonstrating the feasibility of this approach exists in the literature (for an example see: Leman and Naveda (2010)) this would substantially increase the complexity of the analysis framework proposed, thus falling outside the scope of this study.

The movements of the clarinet bell can be much more informative if we use the perspective of the clarinetist as reference frame. In musician's daily practice, bell movements are easier and less disruptive to execute than whole body movements because most of this practice is performed in seating position. This would favor their fixation as reference model for musical matters and, thus, increase their relevance for this study. To focus our analysis only on the clarinet's bell movements, *i.e.* from the musician's perspective, we define the origin of the coordinate system in the interpreter's head. To do so, we calculated the centroid of the markers placed on the clarinetist's head. The origin of the coordinate system was assigned to the musician's head, in order to subtract the influence of translational movements made by participants.

Signals Alignment

In order to carry out the combined analysis of the acoustic and gestural data, it was necessary that all the signals coming from a given recording were perfectly aligned, avoiding a misreading of the obtained results. As previously mentioned, the audio samples from the second session were recorded simultaneously, simulating the recording of a two-part musical excerpt. This process was carried out on two computers, one responsible for the audio and another for the motion capture system. The audio from the microphone was recorded in a dedicated audio interface and forwarded for the Optotrak's ODAU analog-to-digital converter, where it was recorded with a sample rate of 44100 Hz. The ODAU audio is thus synced with the mocap data, although it does note have the proper quality for the audio analysis. This is due to the fact that the signal-to-noise ratio of the ODAU II converter is high, since the equipment is designed to handle amplitude signals around $\pm$ 10 V. In consequence, the noise level in the audio recorded by the ODAU II proved to be significantly higher than expected for the audio signals, making it impossible to detect the temporal events or the proper extraction of any acoustic parameter. The signal recorded by the audio interface had good quality, but the recordings were exported separately and therefore were not aligned. The synchronization between the two signals was accomplished by first finding the lags between the high quality audio and the ODAU audio, and then performing a cross-correlation between the high quality audio and the ODAU audio, thus guaranteeing that the two were perfectly synchronized.

Parametrization

The magnitude of the resultant velocity of each three-dimensional position, estimated by Euclidean distance between two subsequent samples (velocity), was used to parameterize the motion of the clarinet bell, which we named the *velocity profile*, as shown in equation (3.1):

$$v_i = \frac{1}{sf}\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2} \quad (3.1)$$

where x, y e z represent spatial coordinates, and i the sample number. A low-pass, linear-phase Butterworth filter of the sixth order and cutoff frequency of 5 Hz was used for discarding movements of low amplitude, such as those caused by the impact of the fingers on the instrument, or adjustments to the embouchure.

In order to preserve information that may be relevant, such as preparation and finalization gestures, we consider the movement to be analyzed starting at a metrical pulse before the first note and ending at a metric pulse after the end of the last note. The estimation of the tempo of each execution was made from the pulse calculated by the notes on the first bar. The difference between body weights, heights and ages of the participants seem to have influenced the amplitude and speed of movement, resulting in greater variability of absolute speed values across the subjects. In order to minimize the variability related to individual body characteristics and optimize the detection of "gestural signatures" due to temporal details that emerge from the interpretive intentions of the musicians, the amplitude of the *velocity profile* of each performance was normalized by its root mean square value, defined as equation (3.2).

$$v_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{N} v_i^2} \quad (3.2)$$

where v_i is the amplitude of each of the N samples.

Movement data temporal adjustment

Due to the variability of *tempi* of each performance, the velocity curves were adjusted to the same number of samples, using the technique of time warping, as suggested by Wanderley (2002) and Wanderley et al. (2005), which assigns the same tempo for all performances. The mean values for each note onset of all performances was used as a timing model. This procedure aimed at minimizing the misalignment of the velocity curves with the musical structure. The superposition of *veloc-*

ity profiles of two performances by the same musician is shown in Figure 3.4. It can be observed that the misalignment of the two curves intensifies over time, resulting in a total duration difference of ca. 400ms. The bottom panel of Figure 3.4 shows the result of the time warping adjustment. This procedure was applied to each *velocity profile*, allowing performances of different *tempi* to be compared. The instants where these peaks occur vary according to the movements adopted by the interpreter, which makes it difficult to compare executions with different tempos or temporal variations (*rubato*, *accelerando*, etc.).

The first step in the temporal adjustment process is the definition of a temporal model, which will serve as the basis for the normalization of all executions. One option is to create a synthetic model, built from the values extracted from the score. Another possibility is to use the average of the onsets values of all the recordings to create the model. The second option has as an advantage the representation of the instrumentalists' temporal intentions. In this study, we opted for the second possibility. The temporal model used in the time-warping process is defined by a sequence of temporal instants (beginning of each note) calculated as the average of the beginnings of a given note, extracted from all executions. Therefore, this result can be seen as the mean temporal profile of this data set. The velocity profiles are then re-sampled between the values of each time instant provided by this model using cubic spline interpolation. In Figure 3.4 we can verify that the temporal adjustment improves the alignment of the peaks of velocity profiles. This suggests that the movement of the clarinetist follows some type of intrinsic organization that causes the musician to perform certain movements in specific positions of the score, once the velocity profiles were adjusted in relation to the performed notes. One hypothesis raised is that this organization could be dictated by some structural element of the music (*tempi*, rhythms, melodic profile, intensity profile, etc.). We apply this procedure to all recordings in our dataset. This allowed the comparison of recordings made under different temporal constraints which, in the case of this study, are dictated by the leader clarinetist.

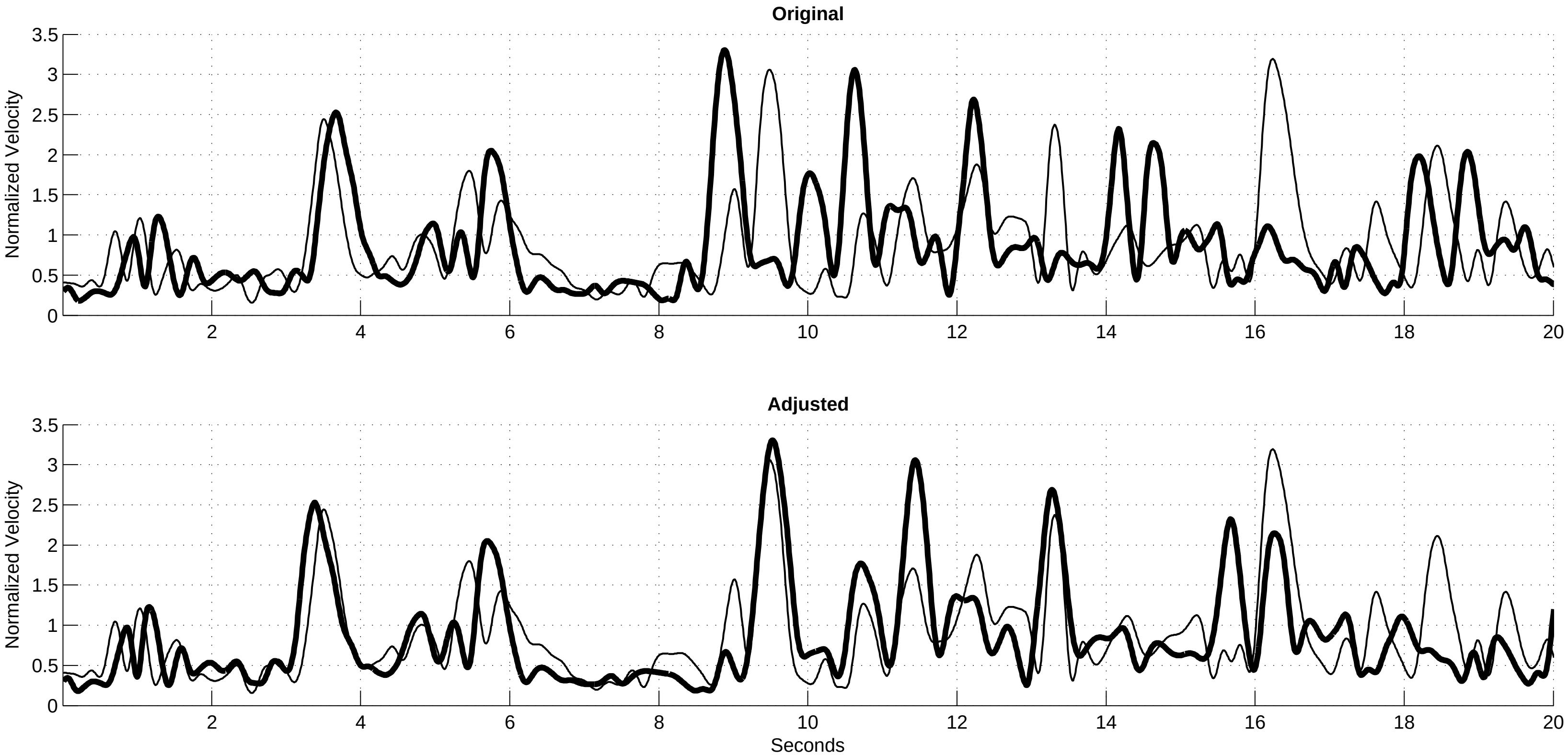

Figure 3.4: Temporal adjustment of *velocity profiles* performed with different *tempi*. The top panel shows the original curves performed by the same musician in two different takes. It is possible to observe the gradual increase of the lag between them. The lower panel shows the same curves after the time warping process, in which the consistency between the two curves can be more easily observed.

4

Synchronization and Consistency

4.1 INTRA- AND INTER-PERFORMER SYNCHRONIZATION IN CLARINET DUOS

As stated in sections 1 and 2.1, the synchronization between musician may depend on several factors which are not yet well understood. In this section we address the question of synchronization in clarinet duos. For this investigation, we use recordings of the Tchaikovsky 5th symphony collected as described in the section 3.1, which contains 24 recordings of solo performances and 36 performances of duets.

Overall synchronization accuracy

Asynchronies were calculated by subtracting leader onsets from follower onsets. Positive values indicate that the follower is playing behind the leader. The overall mean signed asynchrony values were 24 milliseconds, with standard deviation of 111 milliseconds, indicating a tendency for followers to play behind the leaders. This result is fairly obvious if we consider that the followers will try to adjust to leaders' notes **after** they listen to them, which would result in every note being played slightly behind, but it does not seem to be the case. As Figure 4.1 shows, the mean signed asynchrony values for each note in the excerpt tells a different story. As expected, asynchrony values vary widely among notes, suggesting that some notes are harder to follow than others. But, followers' onsets of some notes, such as note 31, were played before the leaders' onsets, suggesting that at those notes musicians tried to predicted the position of the next leader's onset, but failed. Also, higher variability can be observed around at notes 29, 30 and 31, where most mean negative values of asynchrony were measured. Those notes occur in a region with soft dynamics which may limit the comprehension of the recordings and influence the synchronization accuracy. The last five notes also exhibit higher values of asynchrony, which is expect because leaders tended to slightly variate the tempo at the end of phrases. Although signed asynchrony values are as high as 833 milliseconds, note means varied between 216 milliseconds and -68 milliseconds. The highest values of asynchrony occur on the first note, which could be explained by the difficulty in predicting the onset position only by listening the metronome beats played at the beginning of recordings.

Signed asynchronies of isolated notes can be helpful for illustrating tendencies for playing ahead or behind the leader. However, not much can be inferred if we calculate the total mean of all the notes of the excerpt, because the result will tend towards zero. This could be caused by different reasons, for instance, for most of the time, musicians will play **after** the leader, but in some passages they may try to compensate a previous delay and end up playing **ahead**. Also, different musicians or even different takes may influence this result. As we can see in Figure 4.1, of all 36 takes resulting from the interaction between the 6 participants, only notes 1 and 6 have every observed value higher than zero. All the other notes range from negative to positive values, despite their distributions being skewed to one side or the other. Therefore, in this case is hard to define a clear tendency of notes to be ahead or behind, even more a global tendency for all notes together. To accurately assess the effect of asynchronies in the duets we can calculate their absolute values and rely solely on unsigned values. This provides a good way to measure the temporal coordination, although discarding information related to rushes and delays. Figure 4.2 shows unsigned asynchronies for all 63 notes of the Tchaikovsky excerpt. The mean value of unsigned asychronies for the entire excerpt is 77 milliseconds with a standard deviation of 84 milliseconds. Comparison of these results with signed asynchronies ($M = 24.4, SD = 111.9$) reveals how the variety of executions can help to zero out the mean – the lower value for signed asynchrony mean (24.4 ms) does not reveal better adjustment between musicians, since local asynchronies are hidden.

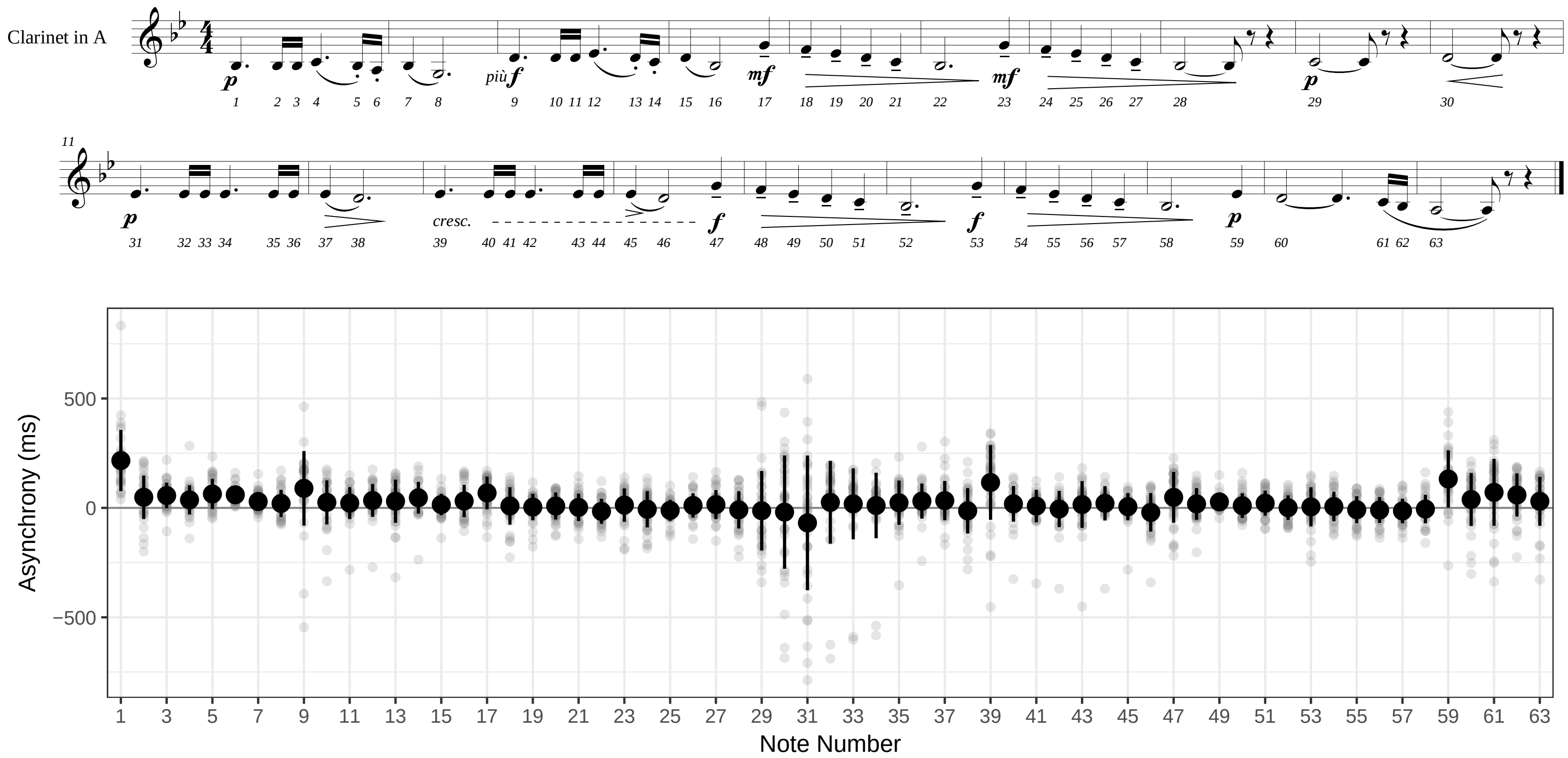

Figure 4.1: Signed asynchronies for all 63 notes of the Tchaikovsky excerpt. Gray dots represent asynchronies values recorded for every note on duet performances. Means are represented by black dots and the error bars indicate one standard deviation of the means.

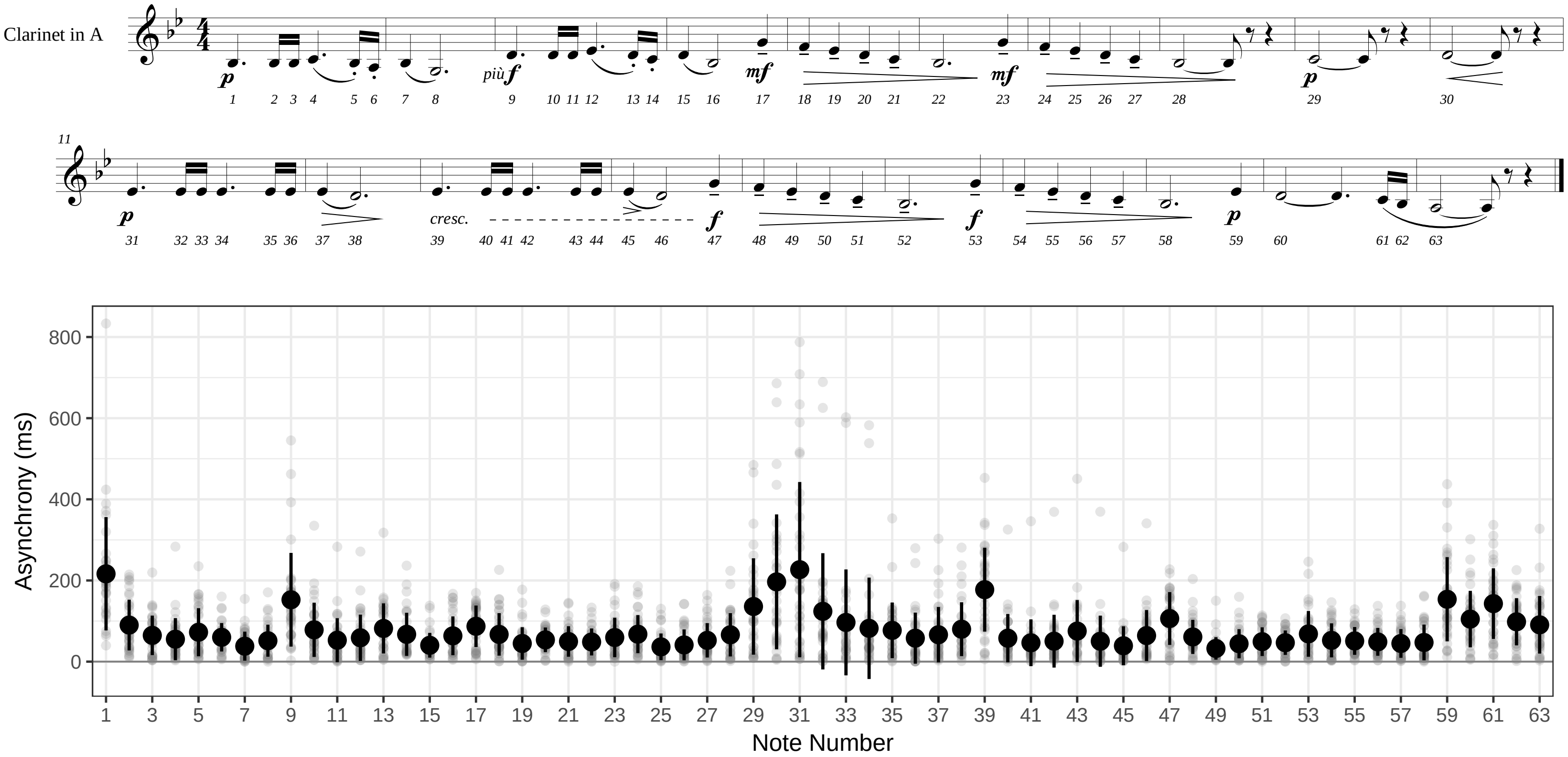

Figure 4.2: Unsigned asynchronies for all 63 notes of the Tchaikovsky excerpt. Gray dots represent asynchronies values recorded for every note on duet performances. Means are represented by black dots and the error bars indicate one standard deviation of the means.

Between performer asynchrony

Musicians that took part in the experiment played following recordings of every other musician, including those made by themselves. This gave us the opportunity to inspect their individual playing characteristics and to examine how they interact with the others. We expect to see different responses to how each participant deal with the dueting task, like differences in the overall synchronization they achieved. Furthermore, we tried to investigate how this could be related to their ease in performing with the others. Individual differences between musicians may arise from different levels of expertise or other idiosyncrasies in their playing styles, like years of experience, musicians from similar performance schools, or trained by the same master, or ensembles with previous performance experience.

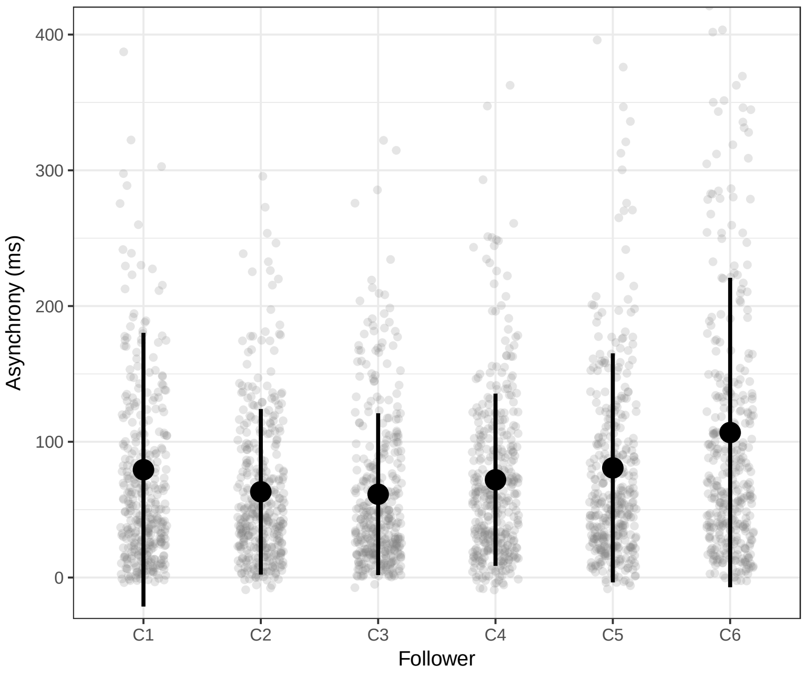

Figure 4.3: Unsigned asynchrony means for 6 clarinetists in the role of follower. Gray dots represent asynchrony values recorded for every note on duet performances. Means are represented by black dots and the error bars indicate one standard deviation of the means.

Figure 4.3 shows the averages for unsigned asynchrony achieved by each participant playing as follower. We can see that each musician have a slightly different asynchrony response as follower, but what stands out the most is the inner variation differences between performers. Some participants have standard deviations as high as 113.9 ms, which is the case of clarinetists **C6** ($M = 106.8, SD = 113.9$) and **C1** ($M = 79.4, SD = 100.8$). Other clarinetists presented lower standard deviation values like **C3** ($M = 61.3, SD = 59.6$) and **C2** ($M = 63.1, SD = 60.9$), which could suggest they had more control over the dueting task. Indeed, by calculating Pearson's correlation coefficient of the average asynchrony values of each follower against their standard deviations we can see that higher values of asynchrony are linked to greater standard deviations, $r(6) = +.91, p < .01$.

We can think about the synchronization in a music ensemble as a moving target point and shoot task, where the success rate would depend on the ability to predict the exact point the target would be at the moment of the shoot. In the case of the musical ensemble, it is very common for leaders to deviate their timing profiles as an expressive resource, sometimes even at live performances, thus making the synchronization accuracy rely mostly on the followers abilities to predict the onset time of the leader's forthcoming notes. From this perspective it seems that followers with lower asynchrony means had more ease in predicting the onset values of forthcoming notes, because the lower standard deviation values would indicate that they try and miss less then the others.

Each follower should also display a different adaptation profile, with some notes or passages being easier to adjust than others. Figure 4.4 shows the superposition of the mean asynchrony values, note by note for each follower. We can see that roughly all musicians have a similar adaptation profile, that is, they tend to have greater difficulty in the same notes, which suggests that a technical or musical issue may be influencing this result. The most difficult musical passages to be synchronized seem to occur near notes 1, 9, 31, 47 and the last five notes. This result could be linked to reduced auditory feedback participants may have had during the performance of those passages, since they occur on regions with lower dynamic markings on the score and they didn't have visual feedback from the leaders to rely on. Nevertheless, some musicians had a more difficult time at some notes than others. For example, clarinetists **C1** and **C6** had more problems synchronizing note 31 than the other clarinetists, which registered almost half the mean asynchrony values on the same note.

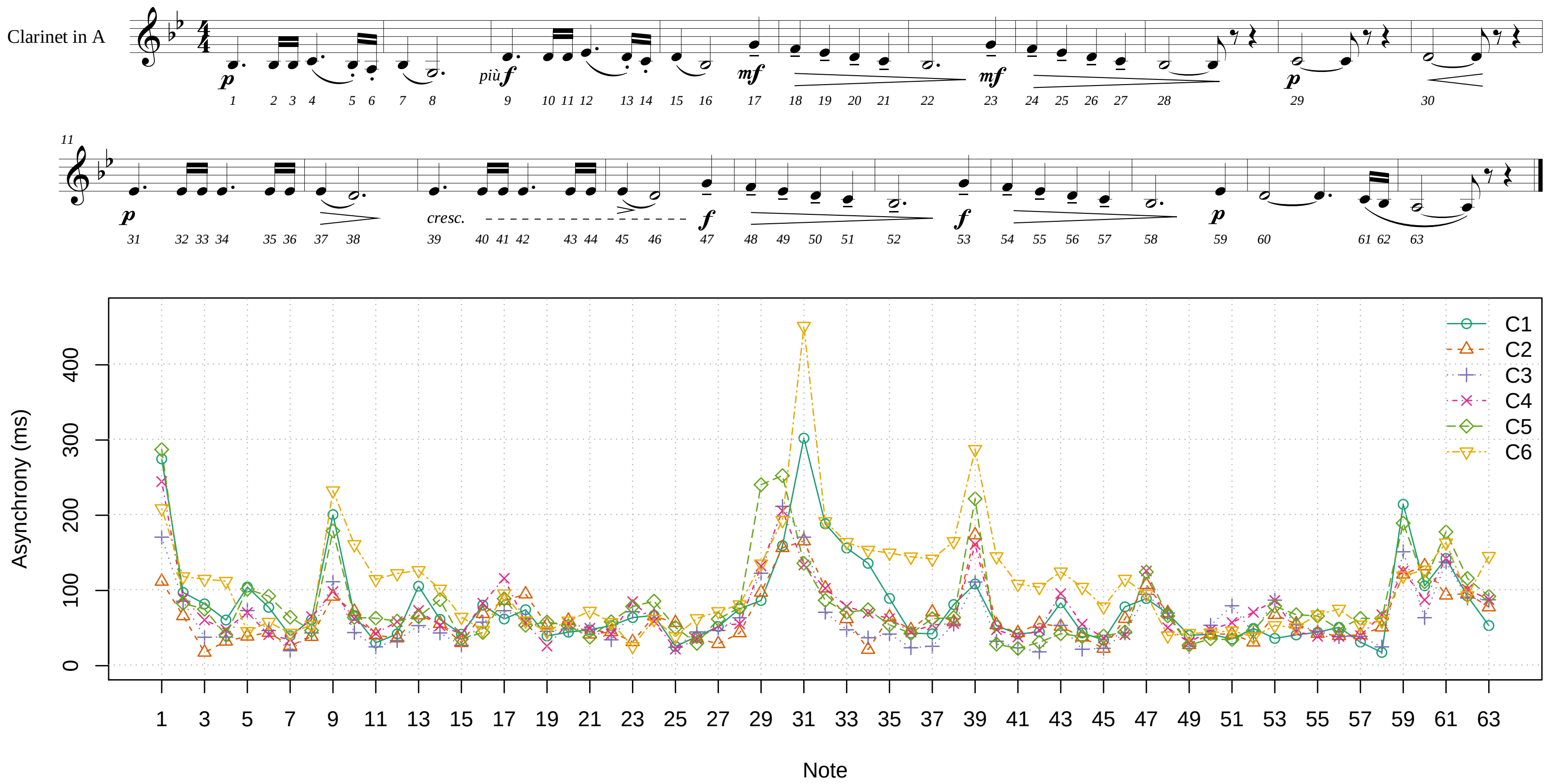

Figure 4.4: Superposition of mean asynchrony values, calculated note by note for each of the six **FOLLOWERS**. Each different follower is displayed with distinct colors, lines and markers.

Leader and follower interactions

As stated before, recordings made in the first session of the experiment were used to represent each participant in the second session, where the task was to follow other participants previous recordings. In the first session participants were instructed to play assuming the role of leader (first clarinet) of the duo. Therefore, we can also investigate the asynchrony induced by each leader, i.e. the amount of asynchrony observed in performances where musicians followed a given leader, and if different leaders can exert different effects on the followers. First, by looking at the overall mean induced by each leader we can see that some of them are harder to follow than others. This is the case of clarinetist **C2** which induces a mean asynchrony of 92 ms ($SD = 93.3$) on its followers, in contrast with clarinetist **C4** that induces an average of 65 ms ($SD = 76.3$) of asynchrony on their co-performers. Figure 4.5 shows the differences between the mean asynchrony induced by each leader.

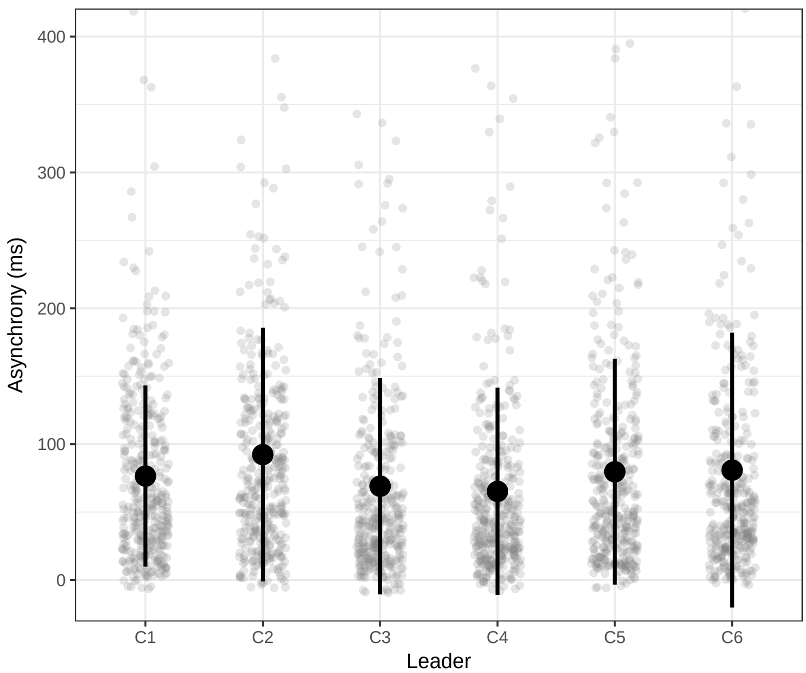

Figure 4.5: Unsigned asynchrony means induced by different clarinetists in the role of leader. Gray dots represent asynchrony values recorded for every note on duet performances. Means are represented by black dots and the error bars indicate one standard deviation of the means.

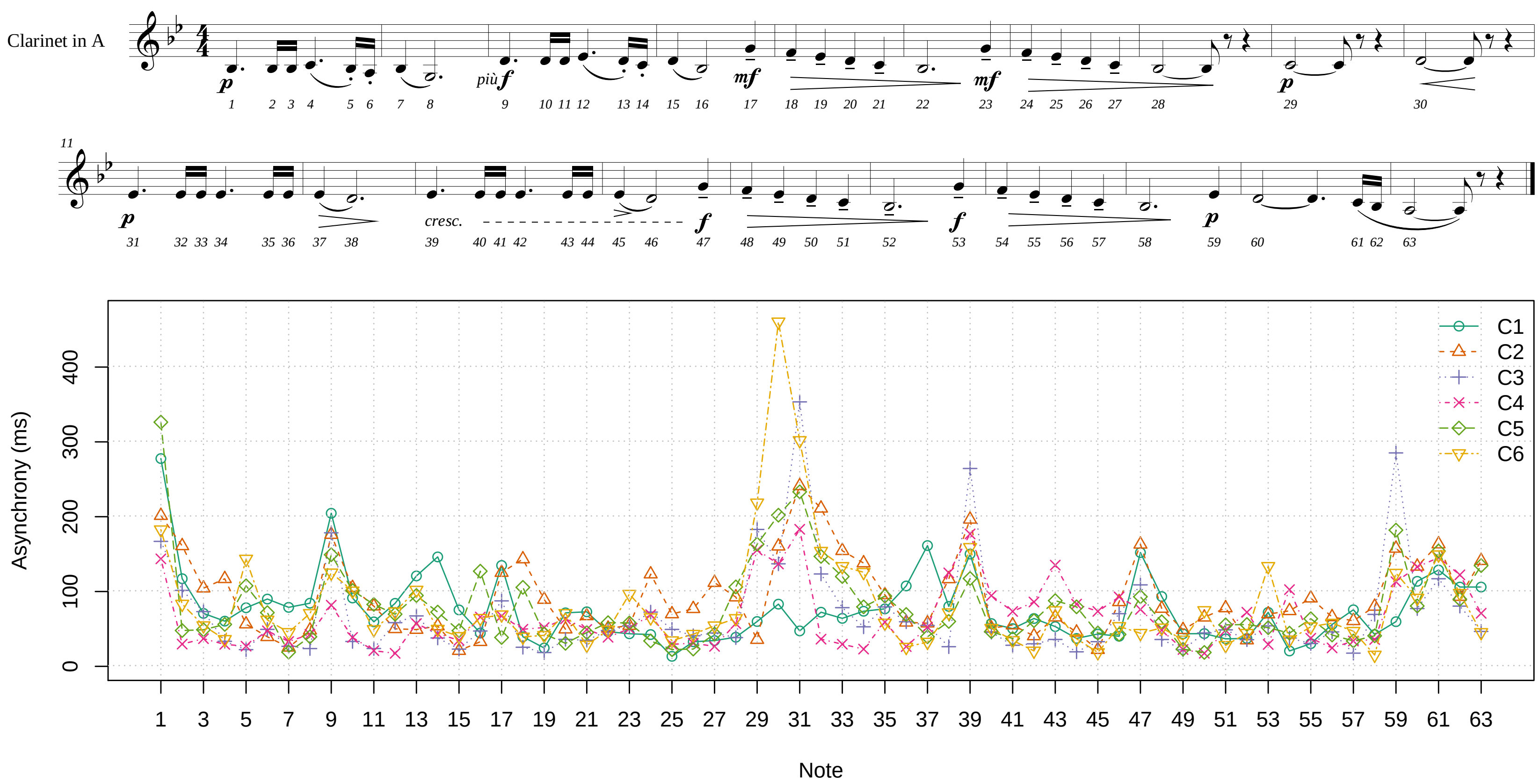

Figure 4.6: Superposition of mean asynchrony values induced by each **LEADER**, calculated note by note. Each different leader is displayed with distinct colors, lines and markers.

As we did with the followers profile, we can also check if there is particular passages or notes where leaders induce more or less asynchrony on co-performers. Figure 4.6 shows how musicians that followed a given leader tended to adjust each note to the leader performance. This result is somehow different from the adaptation profile exhibited by the followers. For instance, clarinetist **C1** induces lower asynchrony values around note 31, breaking the tendency observed on other leaders. This result indicates that the followers could predict better the onset position of notes 30 ($M = 83.1, SD = 83.3$) and 31 ($M = 47.3, SD = 48.4$) with more precision when following the recording made by clarinetist **C1**. On the other hand, on the same note 31, leader clarinetist **C6** induced an average of 301 ms ($SD = 258.6$), suggesting that none of the followers were able to accurately predict the position of this onset.

Self-other

Musicians that took part in the experiment performed following recordings of all other musicians, including those made by themselves. This allowed us to test if the act of following their own previous recordings have an effect on the temporal adjustment of the duo, which is an hypothetical concept of the Common-Coding Theory (HOMMEL et al., 2001). Its fundamental assumption is that actions are coded in terms of the perceptible effects they should generate. Therefore, the observation of the act performed by another person would activate a mental simulation corresponding to that action in our own motor system. Similar motor representations would be activated in the observation and the production of actions. By watching the reproduction of our own previously recorded actions this mechanism would allow us to predict with more precision the outcome of the observed actions. Thus, if internal models are applicable to observed actions, we should be better in predicting results of our own recorded actions than the results of actions performed by others (KNOBLICH; FLACH, 2001). This *self-other* effect was previously observed in several studies where subjects were asked to predict the outcome of actions, like dart throwing (KNOBLICH; FLACH, 2001), handwriting strokes (KNOBLICH et al., 2002), self recognition in point light displays (LOULA et al., 2005) and music-related tasks (KELLER; KNOBLICH; REPP, 2007; LOUREIRO et al., 2011), to cite but a few examples. The cognitive psychology hypotheses behind this concept supports that this behavior is activated through the interaction between action and perception processes, which is referred in literature as *Perception-action Coupling* (NOVEMBRE; KELLER, 2014; LEMAN; NAVEDA, 2010;

KELLER; KNOBLICH; REPP, 2007). In this scenario, action and perception processes would be connected through underlying cognitive functions, common to both systems (JEANNEROD, 2001; WOLFGANG, 1997; SEBANZ; BEKKERING; KNOBLICH, 2006).

The overall mean asynchrony of duets with self-generated performances ($M = 62.1, SD = 58.5$) was indeed 18 milliseconds lower than duets made with other-generated performances ($M = 80.2, SD = 88.4$), which points towards the manifestation of a self-other effect. But before testing this hypothesis we investigated if the synchronization profile along the excerpt, observed on duets with self-generated performances differs from the profile observed on the other duets. Figure 4.7 presents the synchronization profiles for performances with self- and other-generated performances, along the excerpt. Mean values of asynchrony are shown for each note in the excerpt, with self-generated performances (gray dots) and other-generated performances (black dots). As we can see in Figure 4.7, average values of asynchrony for most notes on the self-generated performance condition were lower that the other-generated condition. Most important, both self- and other-generated conditions seems to present a variable adaptation profile, i.e. musicians seem to be trying to actively synchronize the notes instead of simple playing from memory. In the case of the self-generated condition we could expect that musicians would play in the same exact manner as they did on solo performances, resulting in a fixed (or almost fixed) single value of asynchrony for all notes. This result would indicate that in the self-generated condition musicians were not trying to adjust to the leader they were listening to. In that case, the lower values of asynchrony observed would be explained by musicians repeating the same performance as before, but with a somehow fixed delay. The variable adaptation displayed for both conditions in Figure 4.7 suggests that this is not the case, musicians were actively trying to follow all recordings, no matter if the leader was her/himself or other.

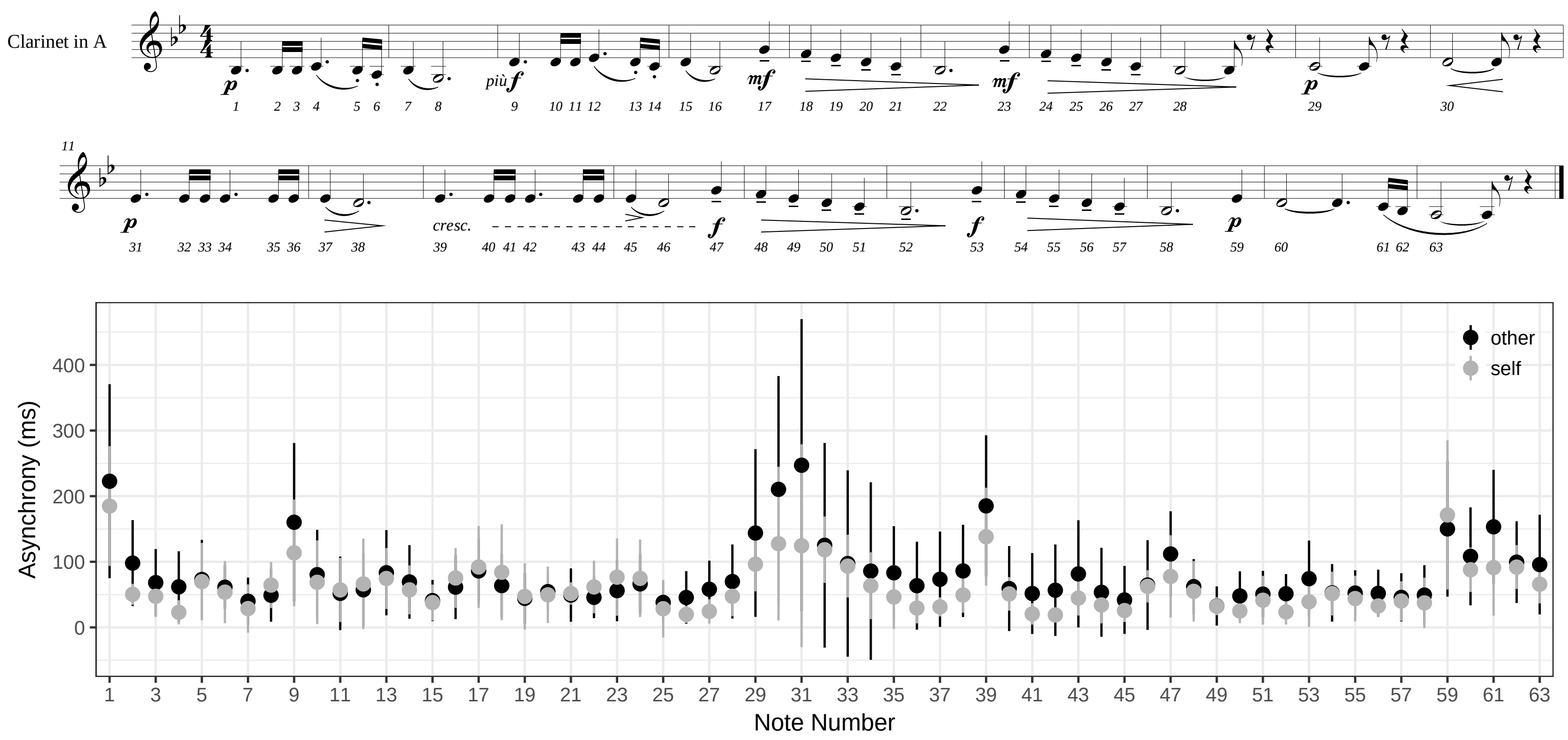

Figure 4.7: Synchronization profiles for performances with self- and other-generated performances, along the excerpt. Mean values of asynchrony are shown for each note in the excerpt, with self-generated performances in gray and other-generated performances in black. Means are represented by dots and the error bars indicate one standard deviation of the means.

How each clarinetist adapt to their colleagues?

The hypothesis that musicians synchronize better with their previously self-generated performances supports that musicians would somehow be able to predict with more precision the outcome of those actions. This implies that the musical information contained in the acoustical stream would be more familiar to them, even if they are not aware of it. Thus, if musicians synchronize better with their own performances, we could also expected that they would respond differently to the different leaders they followed, because some of them may have more familiar playing styles or musical ideas. In order to investigate this possibility we fit a linear regression model with the overall mean as fixed intercept to estimate how much each leader/follower interaction would diverge from it. This approach entails the assumption that the adjustment between musicians would partially depend on some concealed feature of leaders' performances, which would facilitate the follower to achieve lower asynchronies.

As we can see in Table 4.1 most interactions between leaders and followers do not diverge from the overall mean, especially considering the precision (around 6 milliseconds) of the onset detection system used to segment the notes. Interactions with confidence intervals that includes zero within its range cannot be considered to have a strong effect size, even more if their estimates are bellow the threshold of onset detection precision.

After excluding non relevant interactions there seem to be some effect on interactions between follower **C3** and leader **C1** (-26.1 ms, 95% CI -46.4, -5.7), follower **C6** and leader **C1** (42.7 ms, 95% CI = 22.4, 63.1), follower **C6** and leader **C2** (58.2 ms, 95% CI = 37.8, 78.8), follower **C1** and leader **C4** (-33.2 ms, 95% CI -53.5, -12.9), follower **C4** and leader **C4** (-28.5 ms, 95% CI -48.8, -8.2), follower **C6** and leader **C4** (38.2 ms, 95% CI 17.8, 58.5), follower **C6** and leader **C5** (32.7 ms, 95% CI 12.4, 53.1) and follower **C1** and leader **C6** (40.8 ms, 95% CI 20.5, 61.2). Interestingly, some of the lowest asynchrony means observed were not yielded by interactions with self-generated conditions, as predicted by the common-coding theory, such as the interactions between clarinetists **C1** and **C4**. The observed asynchrony when clarinetist **C1** followed himself ($M = 63.8, SD = 54.8$) was around 19 milliseconds higher than when he followed clarinetist **C4** (M = 44, SD = 40.3), and about 33 milliseconds lower than the overall average. However, if we consider all leader-follower combinations, the self-generated performances presented lower asynchrony means, despite these contradictory results for individual interaction pairs. Moreover,

Table 4.1: Summary of linear regression model analisys with the overall mean as fixed intercept and follower-leader interaction as predictor.

		Dependent variable:	
Follower	Leader	Asynchrony	C.I. 95%
C1	C1	−13.4	(−33.7, 6.9)
C2	C1	−12.2	(−32.6, 8.0)
C3	C1	−26.0**	(−46.4, −5.7)
C4	C1	−8.2	(−28.5, 12.1)
C5	C1	12.6	(−7.6, 32.9)
C6	C1	42.7***	(22.4, 63.0)
C1	C2	18.4*	(−1.9, 38.7)
C2	C2	−9.8	(−30.1, 10.5)
C3	C2	0.2	(−20.0, 20.5)
C4	C2	10.3	(−10.0, 30.6)
C5	C2	13.5	(−6.8, 33.8)
C6	C2	58.2***	(37.7, 78.7)
C1	C3	−9.4	(−29.7, 10.9)
C2	C3	−18.9*	(−39.2, 1.4)
C3	C3	−24.3**	(−44.7, −4.0)
C4	C3	−17.0	(−37.3, 3.2)
C5	C3	5.0	(−15.2, 25.3)
C6	C3	15.4	(−4.9, 35.7)
C1	C4	−33.2***	(−53.5, −12.8)
C2	C4	−19.4*	(−39.8, 0.8)
C3	C4	−19.4*	(−39.8, 0.8)
C4	C4	−28.5***	(−48.8, −8.1)
C5	C4	−9.6	(−29.9, 10.6)
C6	C4	38.1***	(17.8, 58.5)
C1	C5	9.8	(−10.4, 30.1)
C2	C5	−13.3	(−33.7, 6.9)
C3	C5	−8.1	(−28.5, 12.1)
C4	C5	−1.7	(−22.0, 18.5)
C5	C5	−4.9	(−25.3, 15.3)
C6	C5	32.7***	(12.4, 53.0)
C1	C6	40.8***	(20.5, 61.1)
C2	C6	−10.6	(−31.0, 9.6)
C3	C6	−17.2*	(−37.6, 3.0)
C4	C6	13.7	(−6.5, 34.0)
C5	C6	4.6	(−15.7, 24.9)
C6	C6	−9.5	(−29.8, 10.8)
Observations	2,267		
R^2	0.065		
Adjusted R^2	0.050		
Residual Std. Error	82.326 (df = 2231)		
F Statistic	4.334*** (df = 36; 2231)		
Note:	*p<0.1; **p<0.05; ***p<0.01		

except for clarinetist **C2**, all participants presented lower mean asynchrony values when following their own performances. Figure 4.8 shows mean asynchrony values for each clarinetist following self- and other-generated performances. Clarinetists are represented by different lines, gray intensities and point shapes.

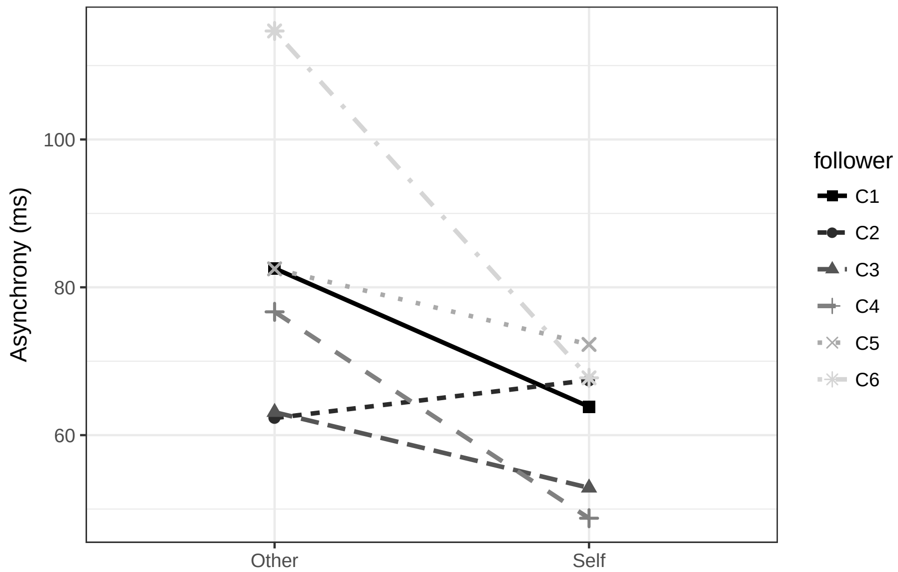

Figure 4.8: Mean asynchrony values for each clarinetist following self- and other-generated performances. Clarinetists are represented by different lines, gray intensities and point shapes.

Comparing models of performer interaction

The effects previously described influence, in different ways, the distribution of asynchrony values observed during the duet task. To estimate the relative importance of those effects, we try to integrate all previous hypotheses into different models and evaluate the contribution of each effect in explaining the observed asynchronies distribution. To do so, we use linear mixed-effects models under the `R` environment (R Core Team, 2017), using the `lme4` package (BATES et al., 2015). A "linear mixed-effects model" describes the linear relationship between response variables (the asynchrony between musicians, in the case of this investigation) and covariates observed during data collection. It is so called

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	80.28***	80.30***	80.29***	80.31***	80.30***
	(5.53)	(6.80)	(3.99)	(7.66)	(9.33)
whoself	−18.12***	−18.14***	−18.13***	−18.15***	−18.14***
	(4.13)	(4.67)	(4.73)	(4.65)	(4.02)
AIC	26062.00	26478.49	26522.44	26468.25	25968.55
BIC	26084.90	26501.39	26545.34	26496.88	26002.91
Log Likelihood	-13027.00	-13235.24	-13257.22	-13229.12	-12978.28
Num. obs.	2267	2267	2267	2267	2267
Num. groups: note	63				63
Var: note (Intercept)	1747.87				1755.21
Var: Residual	5374.74	6883.88	7036.10	6822.05	5084.98
Num. groups: follower		6		6	6
Var: follower (Intercept)		255.74		256.28	260.39
Num. groups: leader			6	6	6
Var: leader (Intercept)			72.99	74.00	78.34

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4.2: Summary comparison of five mixed-effects models using the unsigned asynchrony as the response variable, with the single fixed effect WHO with levels self and other. *P*-values for the fixed effects are calculated from F-test based on Sattethwaite's approximation.

mixed because it has fixed and random effects as predictor variables. In a mixed-effects models one or more covariates have to be categorical, i.e., it has to represent experimental or observational units of the data set, which in our case are considered to be the subjects taking part in the experiment. The fixed portion of a mixed model can be described as a simple linear regression (BATES et al., 2015). In short, mixed-effects models allows us to test if some combination of those fixed effects (attributed to manipulated variables or conditions) and the random effects (which can be attributable to specific sources of random error) would minimize the error on the fitted functions.

For this analysis we fitted five mixed-effects models using the unsigned asynchrony as the response variable, with the single fixed effect WHO with levels *self* and *other*. The random part of the model included the effects FOLLOWER and LEADER, each one with six levels, one for each clarinetist taking part on the experiment and NOTE with 63 levels, one for each note in the excerpt. Models were created in increasing complexity of the random effects structures, from the inclusion of just one random effect to the inclusion of all random effects (LEADER, FOLLOWER and NOTE). To compare those models we use the Akaike information criterion – AIC (SAKAMOTO; ISHIGURO; KITAGAWA, 1986), as a metric for the stepwise selection of the optimal model. To automate the model selection process, we use the package `lmerTest` (KUZNETSOVA; Bruun Brockhoff; Haubo Bojesen Christensen, 2016). According to the documentation of the `step` function from the package `lmerTest`, this process is carried out by first performing a backward elimination of the random part followed by a backward elimination of the fixed part, which is done one effect at a time. Although the `lmerTest` returns p-values our intention is to compare the models to find how much of the random error can be explained by the inclusion of those effects, the lower AIC value obtained by the chosen model along with the effect size of its terms can demonstrate the importance of the results. The comparison of the five mixed-effects models is summarized in Table 4.2. The resulting model incorporated all terms, confirming the reduction of 18 milliseconds in asynchrony values for the self condition ($\beta = -18.1, SE = 4.0, t(2193) = -4.5, p < 0.0001$; 95% CI -26.0 to -10.2). Figure 4.9 presents the results for Fixed effect WHO, showing the mean reduction of unsigned asynchronies for performances in the self-generated condition compared with the other-generated condition. The mean difference between conditions is represented by the circle, the confidence interval is represented by lines around the mean. The horizontal axis represent unsigned asynchronies in milliseconds, zero is the intercept of the model, which is the expected mean value of the response variable (asynchrony) when all coefficients are equal to zero.

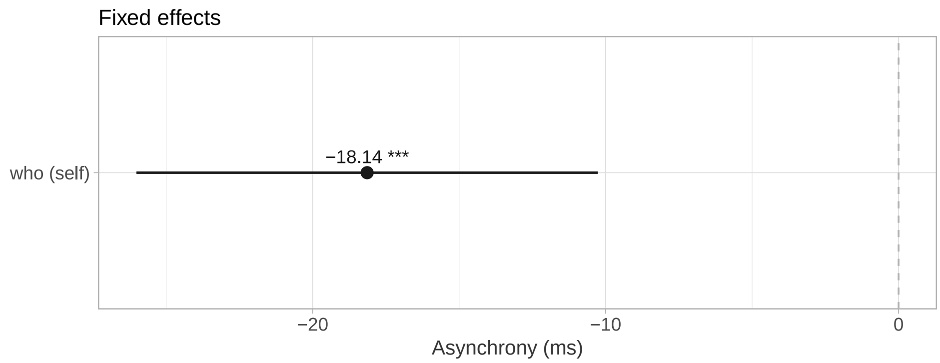

Figure 4.9: Fixed effect WHO, showing the mean reduction of unsigned asynchronies for performances in self-generated condition compared with the other-generated condition. The mean difference between conditions is represented by the circle, the 95% Confidence Interval is represented by lines around the mean. The horizontal axis represent unsigned asynchronies in milliseconds, zero is the intercept of the model, which is the expected mean value of the response variable (asynchrony) when all coefficients are equal to zero.

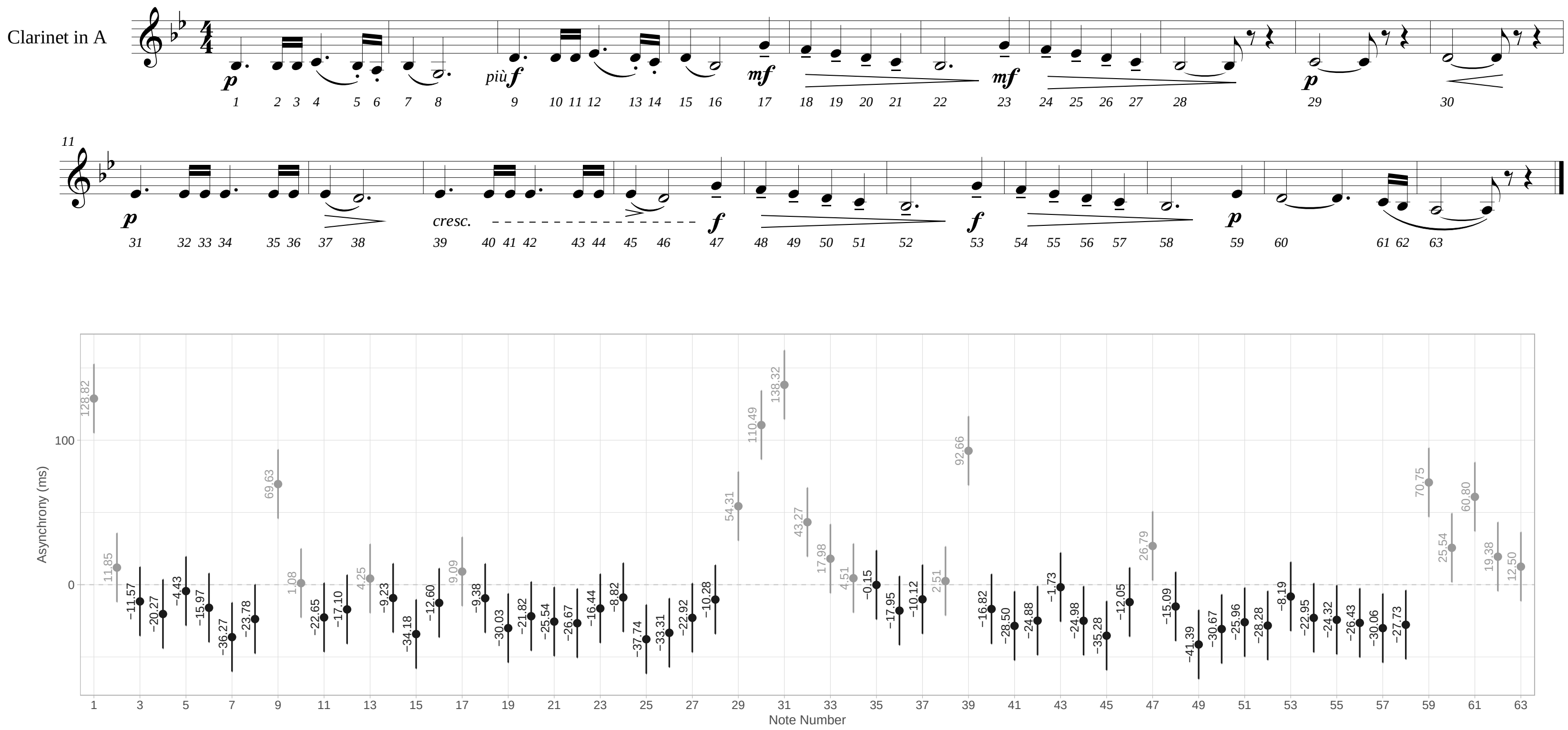

Figure 4.10: Results for the random effect NOTE, indicating the influence of each note on the intercept of the model. Notes that contribute to lower the overall asynchronies are showed in black. Notes in gray contribute to raise overall asynchronies. Vertical bars indicate the 95% Confidence Interval.

The results for the NOTE random effect ($\chi^2_{df(1)} = 501.7, p < 0.0001$) confirms that the influence of the different notes should be taken into account, supporting the previous assumption that some of the notes were more difficult to follow than others. Figure 4.10 shows the results for the random effect NOTE, indicating the influence of each note on the intercept of the model. Results for LEADER random effect (Figure 4.11) indicates that some leaders were more easy to follow then others ($\chi^2_{df(5)} = 19.3, p < 0.0001$). The most successful leader, clarinetist **C4** induced an average reduction of 10 milliseconds in the asynchrony values of their followers, while the least successful leader, clarinetist **C2** induced an average of 12 milliseconds increase in their followers asynchronies. Likewise, results for random-effect FOLLOWER supports that participants responded in different ways to the task ($\chi^2_{df(5)} = 79.5, p < 0.0001$), meaning that some followers were more successful in synchronizing with the leaders than others. The best synchronized follower, clarinetist **C3**, was in average 15 milliseconds more synchronous then the overall mean, while the least synchronized, clarinetist **C6**, was in average 28 milliseconds less synchronous.

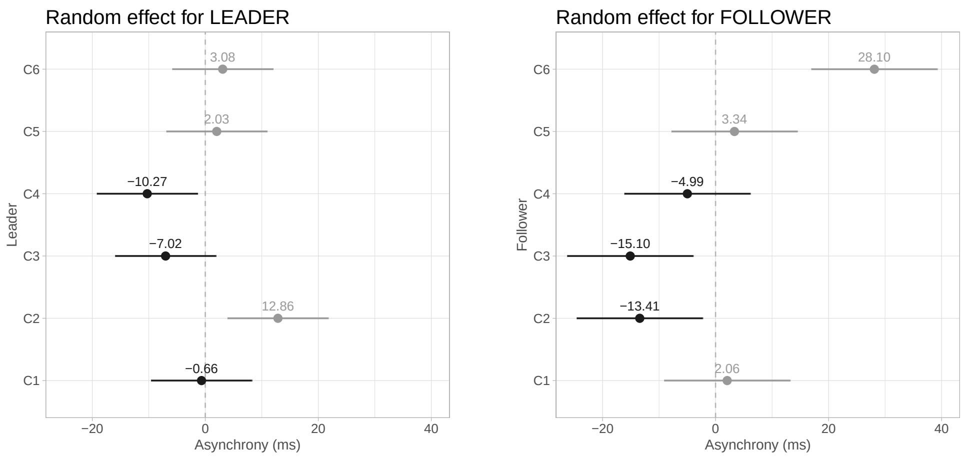

Figure 4.11: Results for random effects LEADER and FOLLOWER showing the average increase/decrease in asynchrony values for interactions with each clarinetist. The estimate for each clarinetist is represented by the circle, the 95% Confidence Interval is represented by lines around the estimate. The horizontal axis represent unsigned asynchronies in milliseconds, zero is the intercept of the model, which is the expected mean value of the response variable (asynchrony) when all coefficients are equal to zero.

4.1.1 *Discussion*

The selected model demonstrates that there was interaction between subjects. In a way, this result is somewhat obvious, as it is to be expected that some musicians would induce greater or lesser asynchronies to their partners performances, and that some notes would be easier to synchronize than others. However, in order to consider the contribution of LEADER and FOLLOWER small effects sizes to the observed asynchrony differences, we need to take into account the perceptual thresholds for detection of auditory onset asynchronies, as well as the adopted methodology for onset detection.

Zera and Green (1993), one of the most prominent studies on the perception of asynchronies, indicates that the threshold for the detection of auditory onset asynchronies can be as low as 1 millisecond. Nevertheless, those tests were performed in a well-controlled laboratory environment, a configuration that distances itself from the complex reality of musical performance. Regarding the detection methodology, as stated in section 3.1.1, the concept of note onset assumed by the system used in this study is based on variations of pitch and RMS values. Note transitions where the frequency varies more than 6% – approximately the size of a semitone, are considered to be a note onset. For repeated notes, i.e subsequent notes of the same pith, our system relies on the small energy gap between the notes for onset detections. In the case of instruments of continuous excitation, the sensation of transitions between notes of the same pitch is created by a small excitation gap. On the clarinet, the instrument used in this study, this is accomplished by the instrumentalist touching her/his tongue on the reed, decreasing its vibration and quickly breaking the air vibration into the tube. This causes a fast and abrupt pitch variation which, allied to the high temporal resolution used in the system, is enough for the detection of the onset on those cases. The criteria used by this automated system guarantees the estimate of onset instances with a resolution of ~6 milliseconds.

The results observed for LEADER and FOLLOWER effects are larger than the human threshold for detection of auditory onset asynchronies, although they are barely outside the resolution of the note segmentation system used. Nonetheless, these results present some evidence that there is a tendency to participants to adjust their interpretation to certain musicians better than others. These results also supports that some musicians seems to have better follower skills than others. Yet, the number of interactions between musicians is not substantial enough for consistently estimate the importance of the observed effects. A number of replications of this experiment have to be done in order to confirm if the size of those effects can be representative of the observed behavior. Apart from that, if we compare, for instance, the LEADER effect size between results for clarinetists **C4** and **C2** the effect increases its importance, with about 22 milliseconds, favoring clarinetist **C4** as the best leader. In this perspective, the difference between those two musicians in the leader role is more likely to be considered relevant, because a 22 milliseconds effect size is further away from the perceptual and technical limits stated before. Within this margin of 22 milliseconds it is more viable to raise questions about which musical parameters are involved in this tendency to induce synchrony or asynchrony. In this context, the immediate question we can raise is why clarinetist **C4** induces lower asynchrony compared with clarinetist **C2**?

If we consider the concept of the Common-Coding Theory (HOMMEL et al., 2001) the follower's ability to predict the outcome of the leader's musical actions would be the most important factor influencing this result. This would mean that the musician in the role of the follower would be able to predict more clearly the leader's intentions if she/he could be able to better understand what would happen in the next note, i.e what kind of temporal, pitch or timbre manipulations the leader would perform. Thus, if the follower is able to better predict the outcome of the leader's musical ideas, the resulting adjustment between them would be better. Another possible explanation would be the resemblance between the leader and the follower musical ideas. This could occur if the follower and the leader had naturally similar musical signatures, so that they would tend to rely their performances in their own musical ideas, as they would do when playing solo, abstaining from interacting – or stop interacting – with the leader for most of the time. Even so, it is hard to believe that participants would not interact at all along the entire performance.

Hence, we envisage two main hypothesis to explain the lower asynchrony values induced by some leaders: I) the follower plays in a very similar fashion to that of the leader with less interaction with the recording; or II) the follower actively interact with the recording and somehow can better predict the musical ideas of the leader. In parallel, this reasoning can be extended to the perspective of the follower. A follower with more experience in ensemble performance would have a superior ability to predict the musical proposition of any leader she/he plays with.

We can discuss the first hypothesis – musicians who have similar musical signatures would have greater ability to adjust, by examining the self interactions, where the musicians play following their own recordings. If they were simply playing from memory, repeating the same rhythmic ideas from their own solo performance, the result of the synchronization would be somewhat invariant, i.e. there would be smaller variations between notes. However, what we observed in the results is the opposite. We observed a note-to-note adjustment, even when they play following their own executions, suggesting they also have difficulty in predicting how certain notes would be played by themselves as leaders, as they do in other-generated interactions. Even when they play with themselves they need to be aware of the musical information being conveyed in order to predict what will happen in the immediate future. We may also expect that the same would happen with musicians who have similar musical signatures, or which came from the same musical schools or share some other level of similarity in the way their musical signatures were constructed. We also would not expect that musicians would simply adjust themselves to the performance of another without actively trying to predict the outcome of the other's actions. The second hypothesis – which indicates that musicians in the role of leader organize their performances in a way that they would be more intelligible, more easier to predict, is more akin to our conceptualization of this interaction process. Another hypothesis would be the influence of the consistency in the execution of rhythmic values. As the musicians who participated in the experiment had access to only one recording of each leader, we can expect that this recording would contain all the relevant information to explain the observed effects, fully encapsulated in the musical information transmitted through the leader performance.

4.1.2 *Conclusion*

This section introduced a discussion about the synchronization in clarinet duos on different performance scenarios. Notably, we tried to observe how each musician adapts her/his performances to previously recorded executions of other players and of their own. We observed that each musician assuming the role of follower responds differently to the task, suggesting that different musicians have different ability levels as follower. In parallel, a similar result was observed for the musicians in the role of leader. Some participants induced greater asynchrony in the performance of their colleagues, while others induced less. We also discussed which elements may influence this behavior. When musicians played following their own performances it was expected that they would achieve a greater synchrony, as observed in the results. But even in those cases there is still an attempt to adjust to the recording, which occurs note by note. This suggests that the familiarity with the recorded performance facilitates the understanding of the musical information being transmitted and increases the predictability of the musical ideas of the leader.

We also discussed the concept of a musical signature. We hypothesized the existence of a musical signature for each performer and that musicians with similar musical signatures would have a better chance to adjust to each other. In this context, the best adjustment case would be when the musician plays following his own previously recorded performance. In the next section we further explore the concept of musical signature, discuss how they may be created and their influence on the adjustment of musical groups. We investigate repetitive rhythmic patterns with various occurrences in the excerpt and try to identify rhythmic proportion signatures that could be occurring. Afterwards, we try to verify if there is any relationship between these rhythmic signatures, their consistency and the results of leader and follower effects that were observed in this section.

4.2 DOES LEADER CONSISTENCY IMPROVES ENSEMBLE SYNCHRONIZATION?

Results of the last section suggested that musicians have different responses to synchronization tasks, which corroborates results of other studies (KELLER; KNOBLICH; REPP, 2007; LOUREIRO et al., 2012). This result is expected, as each participant has to have a particular training experience, which surely leads to idiosyncrasies in their playing style. Nonetheless, this result raises questions about which factors, apart the obvious influence of social-cultural contexts, contributes to make a musician "easier" to play with. A study conducted by Gingras et al. (2011) suggested that the more expressive the judgment about a recording was, the easier was to listeners to recognize the performer. The authors mention the performer consistency in the use of expressive patterns as a relevant factor for the results. In the investigation that follow, we tested recordings from the database to investigate if the rhythmic consistency of performers is somehow related to the asynchrony induced by them when playing in the role of leader. The definition of consistency used in this study can be regarded as a measure of technical precision, which can be represented in terms of similarities in the timing profiles during repeated performances of the same piece (REPP, 1995; WöLLNER; WILLIAMON, 2007).

Prior research suggests that the quality of the fit between interpreters would be linked to the anticipation of the manipulations performed by other members of the ensemble. For example, Keller, Knoblich and Repp (2007) showed that musicians playing on duets achieved greater rhythmic/temporal adjustment when accompanying their own recorded performances, suggesting that musicians can more easily anticipate the manipulations which they are more familiar with. This leads to the hypothesis that the consistency in performing these manipulations could act as a facilitator of the communication between interpreters, meaning that musicians with greater consistency would be better *leaders*. The objective of this investigation is to identify musicians signatures on repeated rhythmic patterns and investigate how they change in order to achieve musical cohesion on duet performances. Furthermore, we try to relate the consistency of different *leaders* to the amount of asynchrony they induce on their partners. In this investigation we used the recordings of Tchaikovsky 5th symphony excerpts, collected in two sessions, as described in Section 3.1.

Signature

As stated in Section 3.1, each participant recorded four *solo* takes as the *Leader* of the duet. They were asked to perform as they would in a real orchestral rehearsal. They were told that their recordings would be used in the second session of the experiment and that other musicians would have to follow those recordings. Our intention was to create a favorable environment for musicians to naturally exercise their expressiveness, in the same way as they would during a real performance. This would allow the idiosyncrasies of each musician, expressed through the manipulations of the different musical parameters, to be revealed. Here the concept of signature is used to describe a group of characteristic marks that serves to set apart or identify an individual, in our case the performer. Our concept of signature goes with the idea of reproducibility. To be considered a signature, a given aspect of some structure (like the qualities of someone's handwriting or the natural features of a mineral) or a behavior (in our case the temporal manipulations performed by the interpreters) must be able to be reproduced in a similar fashion every time they are executed or observed. Furthermore, it must also have its own characteristics that are identifiable by others. The reproducibility which affords our conceptual basis of a signature is therefore linked to consistency in the repetition of the executed patterns. In order for a signature to be unique, individual, its own characteristics must be different from those of the others.

Different studies address the issue of consistency in the execution of a musical excerpt. For example, Repp (1995) assessed pianists consistency through comparisons between their temporal profiles in repeated performances of the same piece. The author suggested that the consistency in musical performances can be seen as a measure of technical precision, taking into account that the musician did not intend to play differently on each repetition. In a more recent study, Wöllner and Williamon (2007) systematically removed the sensory feedback from pianists while they performed repeated performances of the same musical excerpt. They compared the consistency achieved by the pianists on three different conditions, with reduced auditory, visual and kinaesthetic feedback. The authors pointed out that the consistency in musical performance can be measured in terms of similarities of time and intensity profiles, as well as the total duration of the piece in repeated performances.

In general, most studies discussing consistency in musical performance implement some sort of comparison between repeated performances of the same excerpts. In the case of the experiment conducted for our investigation the participants had access to only one recording of the excerpt from each leader clarinetist. Therefore, measuring the consistency by comparing multiple interpretations of the excerpt by one musician is not feasible. Yet, the internal consistency in the repetition of similar rhythmic patterns occurring in the same excerpt could also be regarded as being a manifestation of a signature. The excerpt used in this investigation has rhythmic patterns that are repeated several times, as can be seen in Figure 4.12. Despite the harmonic and melodic differences in each of the occurrences we would expect that the interpretation proposed by the musician for that rhythmic pattern would be executed according to the same reference of note duration proportion. In this context, we hypothesized that greater consistency on the realization of this rhythmic proportion would facilitate the synchronization between the musicians as the follower would more easily predict how the rhythmic pattern would be executed in the next occurrence.

Figure 4.12: Rhythmic pattern chosen for the consistency analysis, with eight occurrences in the Tchaikovsky excerpt. Composed by three rhythmic figures, a dotted quarter-note and two sixteenth-notes, with a total duration of a half-note.

During solo performances the musicians were instructed to follow their own musical ideas. This resulted in small tempo variations between different performances, as well as internal rhythmic variations in each take, as would be expected in a real performance. The purpose for this freedom of interpretation was to encourage musicians to reveal their musical signatures. On the other hand, comparisons of occurrences in the same performance, between performances of one musician, and especially between performances of different musicians became much more difficult. This is because at each occurrence of the rhythmic pattern, note durations would slightly change as a result of the temporal fluctuations employed throughout the performances. This was caused by the slightly different tempo chosen by musicians as well as local expressive timing variations, such as rubatos, ritardandi and accelerandi. In order to compare occurrences of the rhythmic pattern it is essential that all executions were subdued to the same standardized temporal base, so that tempo variations would not interfere with the analysis. This can be done by ignoring their real durations and focusing on their relative durations within the target rhythmical pattern.

A simple method to achieve this, widely described in the literature, is the use of the inter-onset-interval (IOI), the distance between the onsets of two subsequent notes, disregarding any interruption of sound between the two events. Representing note durations as IOI values limit the effect of timing variations occurring before the rhythmic pattern, however it is not sufficient to minimize the problem of comparing the internal rhythmic structure of the pattern, since they might have different durations within the same take due to local expressive timing variation.

In order to compare the realization of the chosen rhythmic pattern, we directed the analysis to the temporal proportion among the constituent notes of the rhythmic pattern, instead of the measured value of their duration. Local rhythmic proportions are less dependent to duration of the total rhythmic figure and are closer to the way musicians subdivide temporal units into shorter rhythmic figures. The ratio between the durations of the rhythmic figures within the rhythmic pattern can reliably represent the individual rhythmic outcome of each occurrence besides being able to be compared to other occurrences of the rhythmic pattern in the same excerpt or in other musicians' performances.

The rhythmic pattern chosen for the analysis is composed by three rhythmic figures, a dotted quarter-note and two sixteenth-notes, with a total duration of a half-note. The resulting rhythm of the pattern is determined by four note onsets, the three onsets of each note of the pattern, plus the onset of the subsequent note. The canonical sub-division of the rhythmic pattern, attributes 75% of the total duration to the dotted quarter-note and 12.5% to each one of the subsequent sixteenth notes. It can be represented by the ratios of its individual figures to a reference duration, in this case the total duration of the pattern. Accordingly, only two ratios of durations are necessary to fully represent the pattern, since a third ratio would be redundant. We chose to consider the dotted-quarter-note proportion related to the total duration of the pattern and the ratio between the duration of the first sixteenth-note and the total duration of both sixteenth-notes.

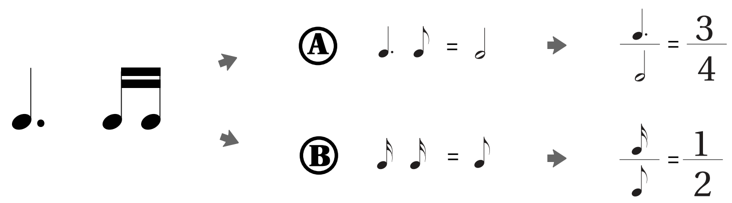

Figure 4.13: Illustration of the two variables used to estimate the consistency of performers: A) the ratio of the IOI (Inter Onset Interval) between the dotted quarter-note to the total duration of the rhythmic pattern (one half-note); and B) the ratio between the first sixteenth-note to the total duration of both sixteenth-notes (an eighth-note).

In summary, two variables were considered in the analysis, as seem in figure 4.13: A) the ratio between the duration of the dotted quarter-note to the total duration of the pattern (one half-note); and B) the ratio between the first sixteenth-note to the total duration of both sixteenth-notes (an eighth-note). These two dimensions allowed us to delineate what can be regarded as an individual performance space, where the performer's musical signature would be manifested. In this individual performance space, each occurrence of the rhythmic pattern is represented by a point in two dimensions, and at each new occurrence of the pattern a new point is marked. As a way to investigate the formation of the rhythmic signatures, the four *solo* performances of each participant were examined in search for idiosyncratic temporal features that could be used to identify interprets. A Multivariate Analysis of Variance (MANOVA) was performed for each pair of musicians using the two variables. By comparing temporal features of different musicians we were able to identify that participants exhibited somehow distinct "rhythmic signatures".

Table 4.3: Resulting MANOVA *p*-values calculated for pairs of clarinetists.

	C2	**C3**	**C4**	**C5**	**C6**
C1	0.15	0.05.	0.57	0.05*	0.00***
C2	-	0.86	0.26	0.79	0.01***
C3	-	-	0.14	0.98	0.01***
C4	-	-	-	0.08.	0.00***
C5	-	-	-	-	0.01***

*0 "***", 0.001 "**", 0.01 "*", 0.05 ".", 0.1 " "*

Table 1 shows *p*-values resulting from MANOVA calculation for each interaction between pairs of clarinetists. Values marked with (*) indicate

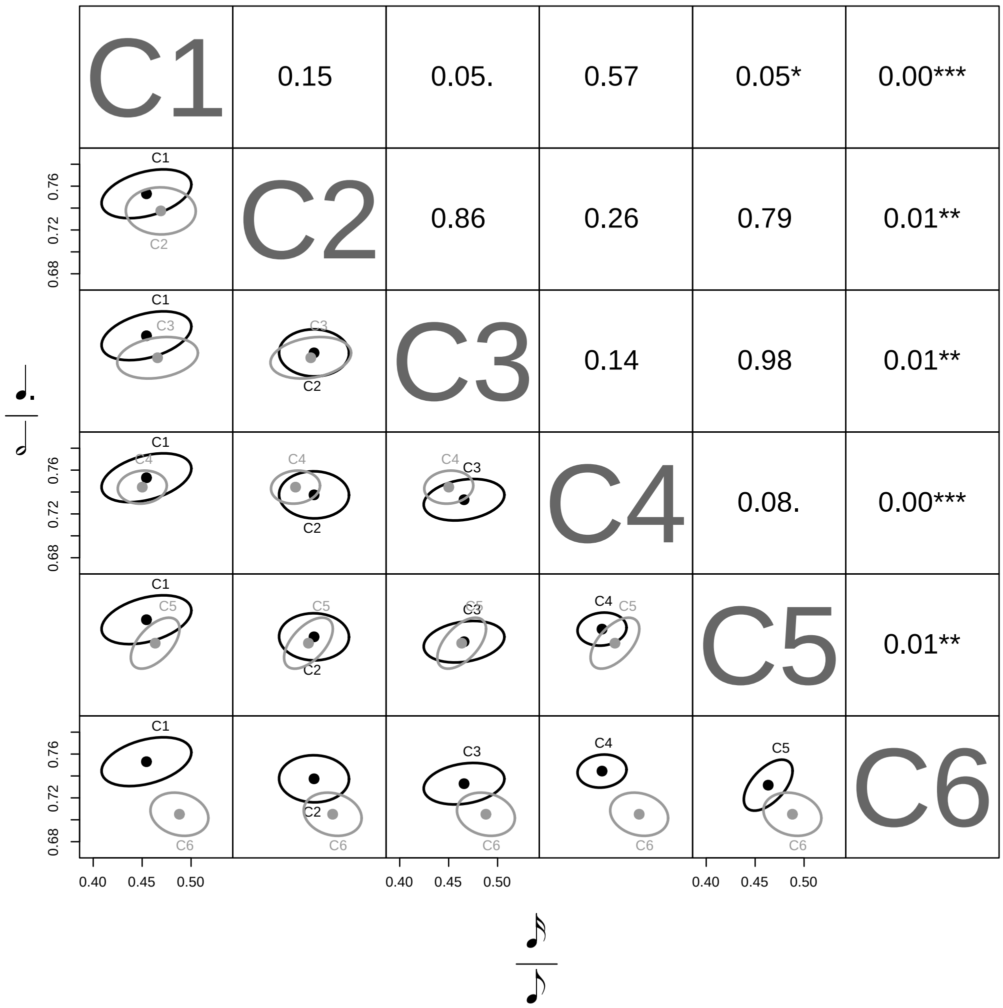

Figure 4.14: Pairwise comparison of individual performance spaces, representing a similarity measure between clarinetists. The upper triangle indicates the *p*-values resulting from a MANOVA calculated between each clarinetist pair. The lower triangle shows the pairwise comparison of ellipses representing a 50% confidence level of each clarinetist distribution.

that clarinetists of that pair perform the division of rhythmic values differently, suggesting they have distinct rhythmic signatures. Pair of clarinetists with values above 0.05 have more similar rhythmic signatures. Figure 4.14 shows the superposition of ellipses calculated as a 50% confidence level of the data distribution for each musician. The rhythmic pattern occurs eight times during the excerpt, therefore each player has 32 observations of each variable (8 occurrences x 4 performances).

Consistency

During the duet performances musicians only listened to one of the recordings made in the first session, the one chosen by each musician to represent her/himself. Thus, all the musical information relevant to the synchronization between leader and follower is expected to be contained in this performance. As discussed earlier, in the individual performance space each point represents one occurrence of the rhythmic pattern on leader performances. The scattering of these points can be used as a metric to calculate the rhythmic consistency of each musician. Musicians who perform the rhythmic pattern always in a similar manner will be considered more consistent, on the other hand, musicians who varies the relative durations of the rhythmic figures at each occurrence will have a lower consistency.

A linear regression analysis was used to investigate the relation between the rhythmic consistency of the leaders and the amount of asynchrony they induce to theirs followers, with asynchrony as the dependent variable and consistency as the predictor. As stated in section 4.1, the asynchrony was calculated by subtracting the onsets of the leader from the onsets of the follower. As mean asynchrony values tended to zero absolute values were used. In order to represent the amount of asynchrony induced by each leader the asynchrony values were averaged across the interactions between each leader with all followers, resulting in six mean values of induced asynchrony, one for each leader. The internal rhythmic consistency of each leader is represented as the standard deviation of the values in the individual performance space, calculated across the eight repetitions of the pattern on the leader solo performance. Incorporating a combination of the standard deviation of both variables did not revealed a relation between asynchronies and consistency, therefore we investigated each one of the variables separately. Figure 4.15 shows the mean leader induced asynchrony versus the standard deviation of rhythmic pattern occurrences. The left panel shows results for the first variable (♩./𝅗𝅥) and the right panel shows the results for the second variable (𝅘𝅥𝅯/♪). Results for the first variable (♩./𝅗𝅥) suggested a relation between the lack of consistency and the asynchrony induced by the leader ($F(1, 4) = 33.26$, $p < .01$, $R2 = .86$), although the results for the second variable (𝅘𝅥𝅯/♪) fail to reveal any connection to the leader consistency ($F(1, 4) = 0.63$, $p > .05$, $R2 = -.07$).

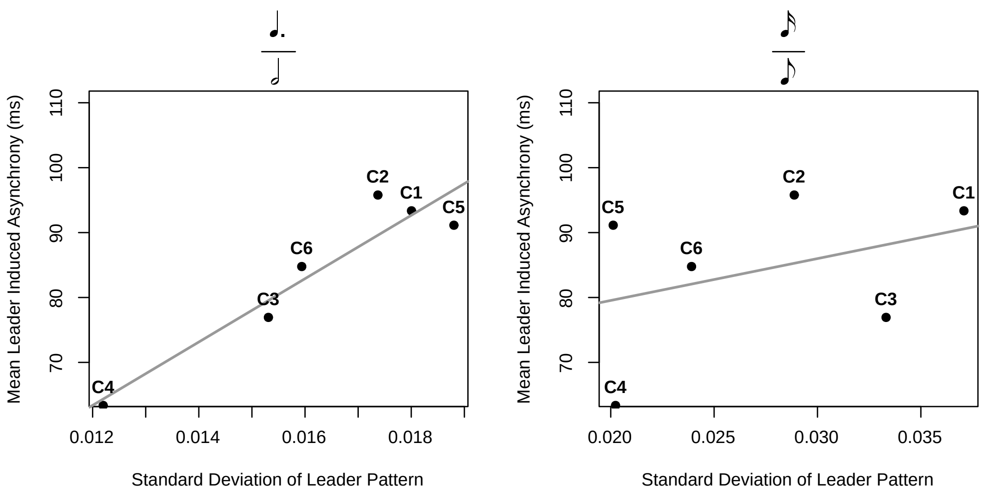

Figure 4.15: Mean leader induced asynchrony versus the standard deviation of rhythmic pattern occurrences. Left panel shows results for the first variable (♩./𝅗𝅥). The right panel shows the results for the second variable (𝅘𝅥𝅯/♪). Asynchrony values are presented in milliseconds. Leaders with higher consistency (smaller standard deviation) induce less asynchrony in the performance of their partners.

This result points towards inconsistency being a predictor of asynchrony in ensembles, but it does not seem to be the case for notes with shorter durations. The question now is why the asynchrony in the first ratio, but not the second, is correlated with the consistency of the leader? One possible explanation would be the duration of the notes, it would be easier for musicians to understand (and hence to predict in future occurrences) the rhythmic subdivision of the first ratio simply because its longer duration would be easier to apprehend, as opposed to the shorter notes as those of the second ratio where the musicians would have difficulty apprehending the subdivision. In other words, the effect of consistency in predicting the rhythmic subdivision performed by the leader would be more noticeable in the longer notes, while in notes with shorter durations followers would have more difficulty predicting the outcome of this subdivision in future occurences, regardless of the leader's consistency in performing that pattern. Indeed Repp, Windsor and Desain (2002) points out several psychological and motor control experiments that suggests qualitative differences in the perception and production of short and long note durations with a discriminatory boundary of about 200 to 300 ms, and hypothesize that "[...] this boundary may reflect a rate limit of a mental clock that generates metrical subdivisions and paces discrete motor actions" (p. 567). The authors also mention evidence supporting the tendency of obligatory grouping of smaller durations, causing "[. . .] interval ratios to be generally perceptually distorted, especially within the range of short durations" (p. 567). Is possible that this effect, in turn, could make musicians naturally shift their attention to notes with larger durations. The distortion in the perception of short durations could explain the difference in the consistency/induced-asynchrony results, which is indeed supported by the independence of the two ratios, which were tested for correlation yielding a fairly poor result: $r(46) = -.14$, $p = .32$. Indeed, values of unsigned asynchrony on the large interval ratio are higher (M = 127.5, SD = 137.9) than the small interval ratio (M = 72.1, SD = 75.7), perhaps mainly due to the local timing variations that occur before the dotted half-note. But higher asynchrony on the dotted half-note doesn't seem to be linked to higher asynchrony on the two sixteenth-notes in the pattern.

Acquaintance to leader rhythmic signatures

Given the observed results suggesting the influence of leader consistency in the asynchrony induced to the followers, we could raise the hypothesis that at each new pattern occurrence, musicians would be more aware of the leader subdivision strategy, making it easier for them to predict its outcome and thus rendering lower asynchrony values. We investigate this hypothesis by testing whether the asynchrony values in leader/follower interactions would decrease in each pattern occurrence. Figure 4.16 shows the average values of unsigned asynchronies of the notes in the rhythmic pattern grouped by occurrence in the score, numbered from 1 to 8. They appear in four pairs, each of them presenting the pattern in sequence without interruptions between them. We can see in Figure 4.16 that the first pattern of each pair has the tendency to assume higher values of asynchrony, with a substantial decrease in the follow up occurrence. This result could implicate that there is a process of acquaintance between leaders and followers taking place over the course of the duet performance, although there is not a clear asynchrony decrease from the first to the eighth occurrence of the pattern. This may be due to the large asynchronies registered in the fifth occurrence, which comes after a region where timing manipulations were abundant. Indeed, by removing the fifth occurrence from the analysis we achieve a reasonable fit in a linear regression with asynchrony as dependent variable and occurrence as the predictor ($b = -4.22$, $F(1, 753) = 10.9$, $p < .001$, $R2 = .014$). This result indicates a 28 ms decrease from the first to the last occurrence, although the size of the effect (4 ms decrease for each occurrence) may still lack significance.

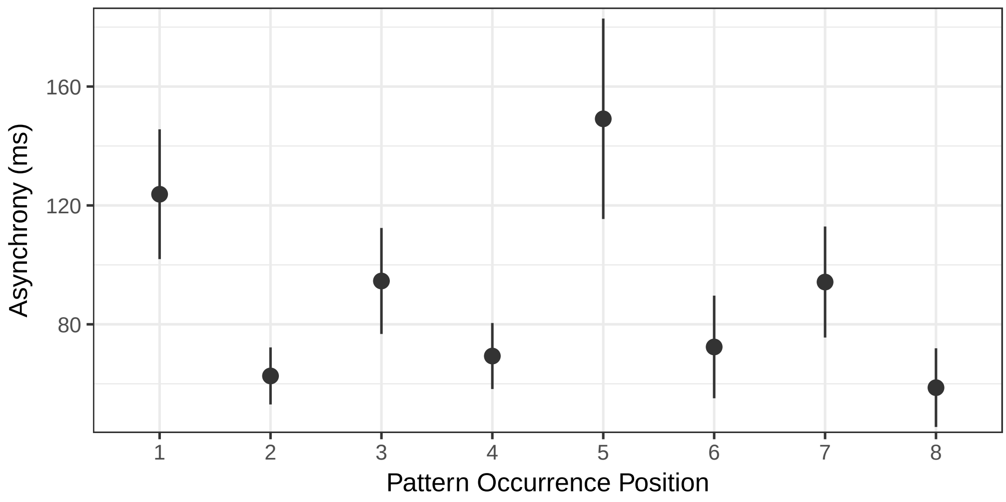

Figure 4.16: Average values of unsigned asynchronies of the notes in the rhythmic pattern grouped by occurrence in the score, numbered from 1 to 8. Vertical bars indicate one standard deviation around the mean.

Adaptation to leader signatures

The hierarchy between leader and follower is part of the culture in orchestral music, although it is normally not imposed, but, on the contrary, is used as a strategy to facilitate the construction of collective performances involving greater number of musicians (ATIK, 1994). Nonetheless, all musicians are expected to exhibit some level of individuality that may be expressed through their performance decisions, whether by rhythmic, timbral or temporal manipulations. We believe that, when those manipulations are performed in a consistent, repetitive way across several performances, such behavior would approach the concept of a signature. When playing solo, all musicians that took part in the experiments described in this investigation exhibited this characteristic in the performance of the chosen rhythmic pattern. Thus, we consider that all participants have somewhat distinctive rhythmic signatures. In this context, musicians playing as followers should voluntarily put themselves in a position where they have to somehow waive their own signatures in order to adjust their performances to the leaders' decisions.

To test this hypothesis we investigated the extent to which musicians sustain their signatures, i.e. the subdivisions of interval ratios, when playing with others. To that end, we compare the subdivisions of the two interval ratios from solo performances with those made as followers in duet performances and conducted a multivariate analysis of variance (MANOVA) with both interval ratios as dependent variables and the condition, solo or duet, as the predictor. We also tested for interaction effects between the condition and the subjects. The results did not revealed any difference between the conditions solo and duet (Pillai Trace = 0.011, $F(2,322) = 1.89$, $p = .15$), nor for the interaction between condition and subjects (Pillai Trace = 0.006, $F(10,646) = 0.20$, $p = .99$). Nonetheless, the results for subject as predictor indicates that even in the duet condition there is a difference between the way each musician perform the interval ratios subdivisions (Pillai Trace = 0.186, $F(10,646) = 6.65$, $p < .0001$), suggesting that musicians do not entirely give up their musical signature, even when following others.

4.2.1 *Conclusion*

The hypothesis raised in this investigation is that **leaders who play more consistently induce less asynchrony** to their followers. This notion seems to be a common sense among performing musicians and was suggested in studies like Repp (1995), although there is still no sufficient data to support this assumption. In this investigation we considered the existence of individual rhythmic signatures on performances of six clarinetists and examined how those rhythmic signatures are affected during ensemble performances. To do so, the rhythmic adjustment between pairs of clarinetists was tested in order to verify how much individual rhythmic signatures were modified when musicians interacted with each other. The results of a MANOVA for pairs of clarinetists suggests the existence of individual rhythmic signatures, which is surprising for a short rhythmic pattern of only three notes. This result also highlighted the different values of consistency in the repetition of rhythmic patterns for each musician. Furthermore, we tested the hypothesis that rhythmic consistency may influence the quality of adjustment between musicians. Results were not conclusive, although they point towards supporting the hypothesis that musicians with higher rhythmic consistency would be easier to follow.

5

Gestural Interactions

5.1 Gestural interactions in ensemble performance

In this section we discuss aspects of gestural interaction in musical ensemble performance, by focusing on gestural responses produced by musicians following recordings of themselves and others. Through the presentation of a case study, we propose that musicians exhibit signature-like gestural patterns while playing solo and that these patterns change due to ensemble interactions. First, we observed that gestural patterns collected across solo performances were consistent enough to allow performers identification, like a signature. Then, we observed that these signatures were disturbed when musicians followed recordings, even without visual contact between co-performers. Moreover, disturbances were smaller when musicians followed recordings made by themselves.

5.1.1 *Leader and follower interactions*

Why some musicians feel more comfortable playing with a particular partner? The answer for this question may rely on technical aspects related to expertise such as rhythmic and tuning precision or either on cultural or social factors, as for example, musicians from similar performance schools, or trained by the same master, or ensembles with previous performance experience. Wöllner and Cañal-Bruland (2010) observed that expertise might also play an important role in this effect. They provided evidence connecting motor expertise and perception of visual cues, by asking string musicians, non-string musicians and non-musicians to identify entries given by a violinist to a string quartet in a series of progressively occluded videos. Hence, it seems that musicians with more expertise in ensembles would tend to have less trouble for understanding the intentions of others, and so accomplishing a more cohesive performance.

So what makes a musician a better follower? Would it be due to resemblance of playing styles or to greater flexibility while playing? A similar question can be extended to leaders. Could it be related to musicians' ability to demonstrate musical intentions in a more comprehensible way? Some studies such Goebl and Palmer (2009) and Fairhurst, Janata and Keller (2014) attempted to investigate this issue. Nonetheless we are still trying to make sense of the underlying processes involved in this interaction. Approaching this issue from the perspectives of leaders and followers might help to reveal the effect that each partner exerts over the other.

In the sequence, we discuss the results of a case study, which draws attention to the effect of auditory cues yielded by leaders on the gestural response of followers in ensemble performance. Specifically, we focus on the different gestural responses produced by musicians while following recordings of others and of their own.

5.1.2 *Parameterization*

We defined a scalar parameter by calculating the normal tangential velocity (speed) using the tri-dimensional coordinates of the clarinet's bell. The resulting curves were filtered using a Butterworth low-pass filter of order 6 and cutoff frequency at 5Hz to suppress high frequency noise in the data. The speed curves contain information about movement distribution along the performance but they have different scales for each musician due to individual differences of height and arm length, which could interfere in the comparison between participants. Therefore, we scaled speed curves by dividing each one by their sum of squares. In addition, when comparing speed curves, regions with low values would appear as being highly similar when in fact this is a consequence of lack of movement. To avoid this unwanted effect we represent the curves in logarithmic scale, further referred as energy curves.

During the recording of solo performances participants were free to perform timing manipulations at their preferred tempo. As overall tempo differences between performances were not relevant to the analysis, we applied a time warping technique, as indicated in Section 3.1.2, in order to be able to compare energy curves of performances of different tempi. The normalization of gestural events in relation to note onset instants rendered a common scale that enables comparison. It was accomplished by re-sampling the energy curves between note onsets according to a common timing model, calculated as the average of each note onset in all executions.

5.1.3 *Results*

First, we tested solo performances for gestural consistency by calculating the standard deviation of all energy curves. The result can be seen as a whole-group gestural consistency measure (Figure 5.1). Higher values represent regions with lower consistency within participants. Lower values represent regions where participants distributed energy in a similar way. As stated before, this measure could be misleading if we had not consider that low variability can also be caused by the lack of energy where musicians do not move at all. Indeed, Wanderley et al. (2005) observed that musicians tended to perform less movements on more technically challenging passages. We observed that participants exhibited similar energy distribution in particular regions, mostly related to energy local maxima, where significant movement took place. Moreover, it is possible that these regions are closely related to musical structure, as the lowest values of standard deviation occurred usually around relevant musical inflections, such as the beginning of important musical phrases (notes 1, 9, 17, 31, 39, 47, 53). We also observed larger values on the final notes of the excerpt, indicating that participants tended to move more differently towards the end of the excerpt. To further explore the individual gestural characteristics of each musician we focus these regions with higher consistency in order to verify whether performers' gestural signatures could be identified. Within these regions of high whole-group gestural consistency, we selected four phrases with the same rhythmic and melodic structure: six descendant phrases of quarter-notes 17 to 22, 23 to 28, 47 to 53, and 53 to 58.

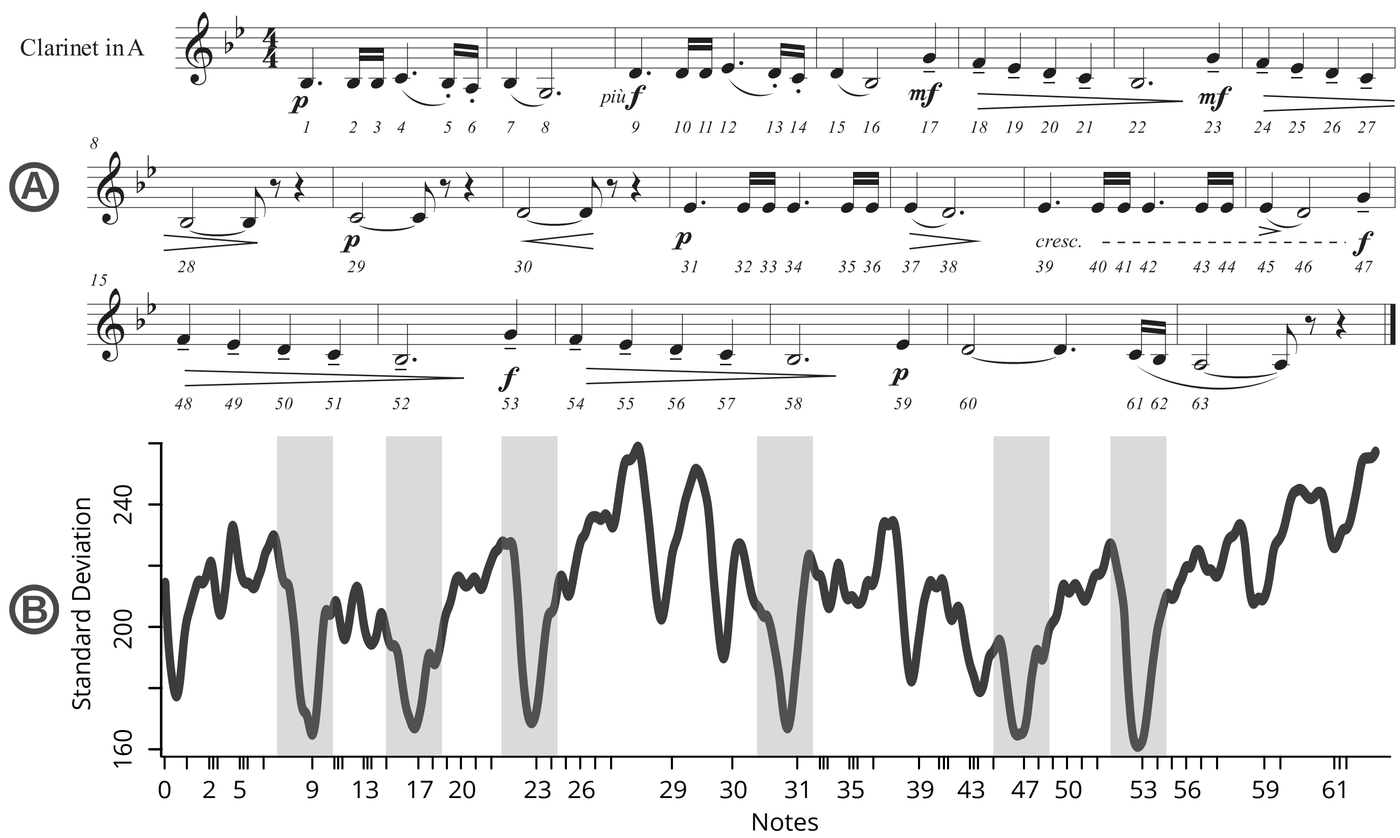

Figure 5.1: Opening bars of the first movement of Symphony No. 5 in E minor, Op. 64 by Pyotr Ilyich Tchaikovsky (Panel A). Standard deviation of solo speed curves and consistency regions highlighted in grey (Panel B).

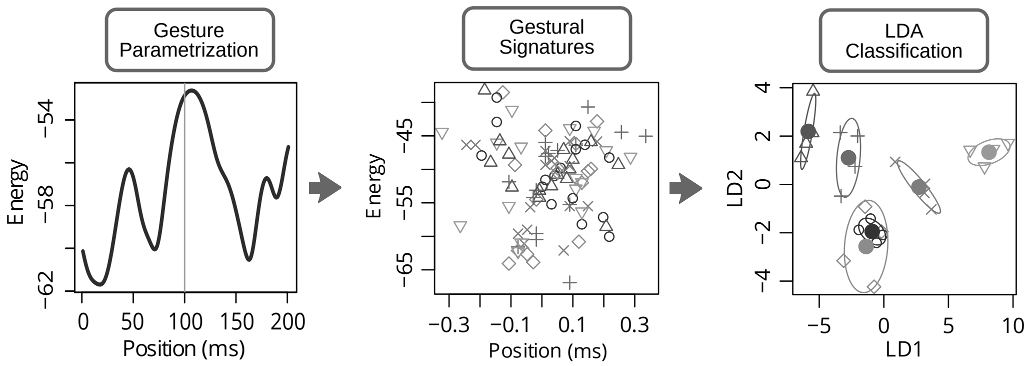

Figure 5.2: Overview of the gesture analysis process. Gesture parameterization (LEFT), using two variables, the position and the value of the peak; signature identification (CENTER), representation of each performance as a single point in an 8-dimensional feature space; LDA classification (RIGHT), search for a linear combination of features that characterizes each performer.

On initial visual inspection it was possible to observe some important differences in the movements performed on the vicinity of the four chosen phrases: (1) some musicians tended to make larger, slower movements, while other musicians tended to make shorter, faster movements; (2) some musicians performed movements before the first note of the phrase, others during the first note, and others after the first note. Taking into account the observed characteristics we applied a simple parameterization consisting of two variables, the position and the value of the nearest peak registered in the energy curve around the first note of the phrase (Figure 5.2, left and middle panels).

Linear Discriminant Analysis (LDA) was used to test whether the observed individual characteristics would be sufficient to separate the musicians (Figure 5.2, right panel). Each of the four solo takes recorded in the first session were represented as a single point in an 8-dimensional feature space, corresponding to the position and energy values for each phrase. This arrangement was chosen to account for inherent musical differences between the phrases, which could induce different gestural responses. The LDA model correctly classified 100% of solo performances, even though the simplified parameterization applied discards part of the information contained in the energy curves. The high accuracy of the identification of musicians is an indicative of how participants maintained their gestural signatures along different executions of the excerpt.

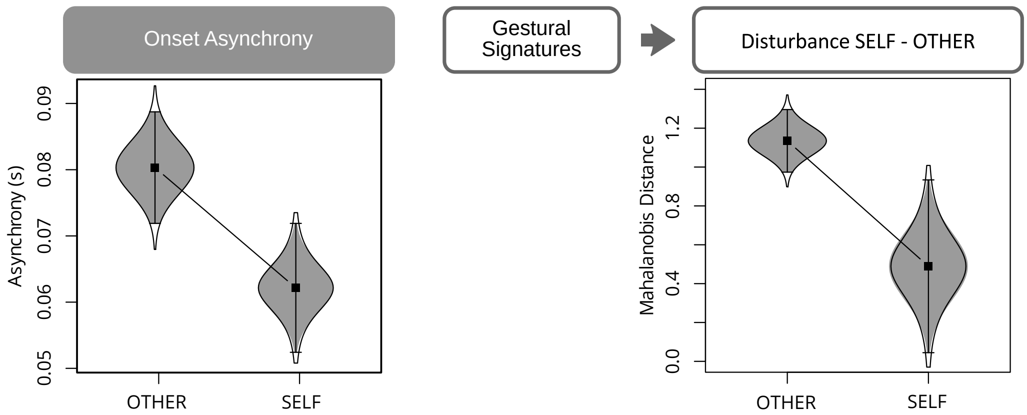

Figure 5.3: Asynchrony and gestural signatures disturbance during self-self and self-other interactions.

Next, we examined movement data collected during the second recording session, when participants performed following recordings of themselves and of the others, aiming at verifying if any disruption on gestural signatures due to such interaction could be detected. We adopted the same parameterization of movement data used with solo performances. Data collected on the first session (*solo* recordings) was used to estimate an average gestural signature for each musician. We proceeded by calculating, for each musician, distances between the average solo gestural signature and the gestural signatures collected during duet performances. This would measure how much musicians moved away from their original gestural signatures when playing as second clarinetist, referred here as "disturbance". Due to scale differences between the two dimensions, Mahalanobis distance (MAHALANOBIS, 1936) was used. We considered the hypothesis that musicians would tend to keep their original gestural signatures when following their own recordings, but not when following others. Disturbance values were first log-transformed to remove skewness and then averaged across interactions between leaders and followers. Mean disturbance of self-self duets (M = 0.48; SD = 0.42) were lower than self-other duets (M = 1.13, SD = 0.43), $t(34) = 3.35$, $p < .01$, 95% CI [0.25, 1.03], suggesting that participants were more likely to preserve their solo gestural signatures when duetting with themselves than when duetting with others (Figure 5.3, right panel).

Note synchronization accuracy was also investigated by subtracting onsets of the follower from the onsets of the leader. Mean asynchrony values tended to zero, therefore, absolute values were used in the analysis. Around 25% of the notes had values of absolute mean asynchrony higher than 100 ms. The majority of those notes were either the first or the last five notes of the excerpt. Errors on the first note can be explained by the difficulty of predicting the exact moment where the onset would occur, despite the presence of metronome beats at the beginning of leader recordings. Synchronization of phrase endings tends to be more difficult due to timing variations performed by leaders, as most musicians do. Both cases may be related to the lack of visual interaction between leader and follower during recording sessions. Asynchrony values were averaged across interactions between leaders and followers and Welch's approximation was used to account for unequal variances. The overall asynchrony for the whole group was 77 ms (SD = 84 ms). Mean asynchrony values for conditions OTHER and SELF were 80 ms (SD = 9.27 ms) and 62 ms (SD = 22.55 ms) respectively. Asynchrony values were 18 ms lower when participants followed themselves, $t(19.199) = 3.24$, $p < .01$, 95% CI [6.45, 29.85] (Figure 5.3, left panel).

5.1.4 *Discussion*

The musical relevance of regions where gestural consistency was observed in solo performances, suggests that temporal energy allocation of movement may be related to internal representations of expressive intentions. It is possible that those movements act as preparatory gestures, perhaps condensing an undetermined amount of mixed information into one single gesture, which in turn would function as a mental representation of the desired acoustical outcome. This would facilitate the musician to refer to an eventual inner "musical expression library" containing predefined solutions for that particular set of musical actions. Indeed, Leman and Naveda (2010), whilst analyzing gestures of dancers, pointed out that a pattern of repetitive gestures could act as a mental representation, a spatial temporal reference frame for dance patterns and/or could be related to the motor domain, where it would act as a reference for motor activity in response to auditory input.

Higher variability of energy curves, observed at the end of the excerpt, indicated that participants tended to move with less consistency in comparison to other parts of the score. In turn, such decrease of gestural consistency suggests that it could be related to the higher timing variability observed. Moreover, the challenge of note synchronization, as well as of expressive coordination in the absence of visual information might also explain the gestural disturbance observed in these passages. The choice of high consistency regions to investigate individual gestural characteristics was not arbitrary. We aimed at envisaging multiple hierarchical levels of discrimination from the more general, such as musical style, or musical school, to the more individual, such as the ensemble, or the performer. Perhaps, gesture disturbances could be seen as a reflection of how far distant two musicians are situated across these levels - would musicians with similar gestures synchronize better?

It is clear that the duetting task influenced the gestural signature of the participants, by changing both temporal position and energy distribution of the gestures relative to the score. Higher values of energy ensures that this effect is not related to eventual decrease in the amount of movement. They seem to be related to musicians' attempt to adjust to the musical intentions proposed by the leader. As there was no visual information exchanged during the recordings, it is not possible to determine if this pattern change would be caused by an impulse to produce visual cues to the partner. The asynchrony results suggested that musicians tend to better adjust to their own recorded performances comparing to recordings made by others. A similar result was observed for movement data, as the disturbance of gestural signatures of participants was smaller when they followed their own recordings. The influence of self-generated actions in the outcome of musical tasks was observed in previous studies, which also indicated that self-generated actions usually achieve better accuracy in the results of the performed task (KELLER; APPEL, 2010; KELLER; KNOBLICH; REPP, 2007; WÖLLNER, 2012). Keller, Knoblich and Repp (2007) also suggests that the recognition of musical actions are related to motor components, by means of internal representations and motor imagery. Thus, the synchronization of musicians would be achieved by each interpreter upon internal simulation of the actions of other members of the group, based initially on how they would perform the same excerpt. Therefore, when duetting with themselves, musicians would associate the musical actions of the leader to self-generated actions, which, in turn, would facilitate synchronization.

When musicians play in ensembles they face a complex task, which is to align their individual performance choices to those of the co-performers. Consequently, the flexibility of a musician should play a significant role in the overall adjustment of the ensemble. The disturbance observed during the duets could be a reflection of this adjustment and may be related to the accompanying skills of the follower. Further investigation in that direction could reveal how gestural disturbance could relate to measured synchrony, and therefore to musicians accompanying skills.

An intriguing hypothesis would be that musicians have predefined motor schemes that are triggered when they perform. Those motor schemes would be built along years of training, under the influence of numerous (not yet quantifiable) variables, which would also be influenced by the interaction with other musicians. One may say that this motor scheme would function as a reference table to a collection of interpretation strategies encapsulated in simple gestural events. Thus, when performing a certain musical action, musicians would use this motor scheme as a way to gain access to those predefined strategies. When facing a situation that conflicts with their own predefined musical settings, musicians would have reduced control of this motor schema - disturbance on gestural signatures would reflect the challenge posed by the musical interaction demanded by the task.

5.1.5 *Conclusion*

In solo musical performances, recurrent patterns can be easily observed in the way musicians manipulate audio features. This recurrence could be described as a “musical signature” which is a representation of the individual interpretation choices made by the performer. We were also able to observe that musicians presented distinctive, highly consistent, patterns of body movement while performing. When comparing performances of six different musicians, regions of inter-performer commonality, i.e regions where all musicians performed similar gestures, were detected. Those regions tended to occur in positions of great musical significance, which strongly suggests a connection of such behavior to the musical structure. Furthermore, inside those regions of commonality, we were able to detect individual characteristics, a gestural signature of each performer. Duet performances were investigated in search for disturbances of gestural signatures of the followers. Results indicated that gesture signatures were less disturbed when musicians followed their own recordings, as compared to the situation when they followed recordings made by others. Moreover, a similar effect was observed on audio asynchrony, corroborating findings from other studies. To conclude, we believe that further investigation on gestural disturbance and its possible association to audio synchronization could give new insights on the underlying mechanisms involved in ensemble performance interaction.

5.2 INFLUENCE OF EXPRESSIVE COUPLING IN ENSEMBLE PERFORMANCE ON MUSICIANS BODY MOVEMENT

This section approaches the gestural adjustment between *leader* and *follower*. We tested the hypothesis that while playing on ensemble, the "gestural signatures" of a musician suffers the influence of the "gestural signatures" of the colleagues. To test this hypothesis, we try to demonstrate that: (i) the kinematic data contains sufficient information to identify "gestural signatures" of musicians; and (ii) the "gestural signatures" of different musicians could be influenced by different situations of interpretation, for example when following the expressive intentions of the *leader*. Evidence of modifications were found in the "gestural signatures" of the *followers* when they followed different *leaders*, even when there was no eye contact between them.

Gesture and music

Several empirical studies in music performance have shown evidences that musicians manipulate note durations, articulations, intensity, pitch and timbre, in order to convey musical intentions of a particular interpretation (GABRIELSSON, 2003). Notable differences may arise between interpretations of distinct performers or even between the same performer in different situations (PALMER, 1997). Constancy on such manipulations may be acknowledged as a style or a signature of the interpreter. It is also well known that body movements in music performance also communicate interpretative intentions. In recent years, great efforts have been devoted to the study of these movements. Cadoz, Wanderley et al. (2000) proposed to differentiate body movements directly related to the production of sound (instrumental gestures) from those that are not (ancillary gestures), suggesting that the latter present tighter relations to the performer's expressive intentions. Attempts have been made to characterize and quantify physical gestures involved in musical performance, in search for their musical significance (WANDERLEY et al., 2005; RASAMIMANANA, 2012; DESMET et al., 2012). In order to identify how the information contained in body movements relates to the music structure and consequently to the musician's intention, some authors tried to create segmentation models of this gestural data (TEIXEIRA et al., 2015; CARAMIAUX; WANDERLEY; BEVILACQUA, 2012). Teixeira et al. (2015) investigated the musical significance of clarinetists' gestures in performances of excerpts from classical and romantic repertoire. They were able to detect high recurrence of movement activity correlated to relevant harmonic and melodic changes, which they considered as evidence that musical significance is expressed in musician's body movement.

Ensemble performance

In instrumental ensemble performances, musicians have to coordinate their actions in order to converge to musical cohesion, which enables the accomplishment of a consistent performance where not only the notes are synchronized, but also the musical ideas are coordinated. To do so, musicians have to anticipate the expressive manipulations of the notes played by other members of the group. The burden of this coordination is shared among all musicians engaged in the musical task, either playing as a leader (serving as reference for other players, such as a conductor, a *spalla*, or *Clarinet I*), or as a follower of the musical interpretation proposed by the leader. As pointed out by Gabrielsson (2003), the goal of the movement performed by a musician, in addition to giving relevant information for the coordination with others, may also be used for communicating expressive intentions, which provide information about the artist's personality or simply entertain the audience.

Goals of the present investigation

Even though ancillary movements have important roles in the transmission of expressiveness and in the synchronization of ensemble performances, the precise way in which instrumentalists adjust their movements when playing with others is still an open research question. The present section is aimed at an empirical exploration of this issue. More precisely, we sought to determine whether the body movement of musicians contains information related to the interpretative intentions in a performance. To test this hypothesis, we attempt to demonstrate: (i) that body movement data contain sufficient information to identify "gestural signatures" of musicians from recurrent kinematic patterns; and (ii) whether "gestural signatures" of different musicians could be influenced by different interpretive situations, for example in instrumental duet performances, where musicians have to follow the musical conception of the leader.

Although musicians take advantage of visual information conveyed by body movements of other members of the ensemble, in order to improve their synchronization and overall musical coordination, they are able to follow other musicians solely by listening to what they play, without any visual contact. This facilitates the methodological design for approaching the question proposed by this investigation: does musicians change their gestural signature when playing with others?

5.2.1 *Results*

Musical gesture recognition

Results of previous studies suggests that musicians have a “musical signature”. This signature can be observed in acoustic parameters that describe characteristics of tempi, timbre and articulation of notes. For instance, a significant decrease in mean asynchrony between notes, measured over 4 subsequent takes, was observed in Loureiro et al. (2012), suggesting that musicians have the ability to quickly learn to predict the expressive intentions of their partners, which may indicate evidence of interpretive coupling in ensemble performance. As observed by different studies (WANDERLEY, 2002; WANDERLEY et al., 2005; NUSSECK; WANDERLEY, 2009), the consistency in gestural patterns exhibited by musicians performing similar musical content suggests that they might also present “gesture signatures”. In fact, “gesture signatures” in everyday task performances were demonstrated by several studies, such as Farella et al. (2006) and Loula et al. (2005). In this investigation, the existence of individual “gestural signatures” was evaluated with pattern recognition techniques applied to the kinematic data extracted from solo performances of clarinetists. We use recordings of the Stravinsky excerpt extracted from the ballet Petrushka – the Quatrième tableau Nº 100, first three bars, collected as described in the section 3.1. For the analysis we used the instantaneous velocity of the centroid of the markers attached to the bell of the clarinet, and applied the time warping technique described in Section 3.1.2 to minimize the misalignment of the signals. This causes all velocity curves to have the same number of samples, around 1500 at the 100Hz frame rate used for the motion capture. Thus, we consider each of these samples as one dimension in the data space and applied Principal Component Analysis (PCA) to further prepare the dataset for the K-means Cluster Analysis, applied to identify the 6 players.

Principal Component Analysis was able to explain over 90% of the

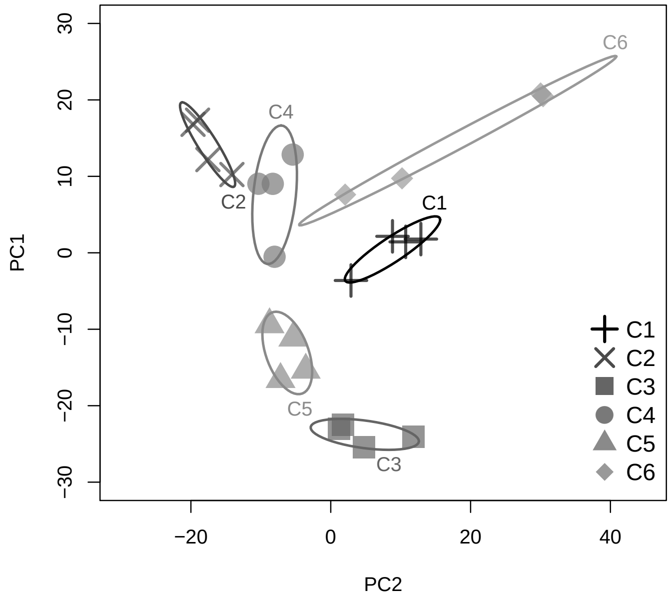

Figure 5.4: K-means clustering of the solo performances, represented in a two-dimension subspace composed by the two first PCs. The velocity curves of the four solo executions of each clarinetist are shown with points in different shapes for each subject. The ellipses show the result of the k-means algorithm when six classes are required.

dataset variance with 13 principal components. Note that the goal of PCA here is not to find a reduced space for the representation of the data, even though this process led to a reduction of dimensionality. Actually, reducing the number of explanatory components is a necessary step for avoiding the "high dimensional, low sample size data" problem (QIAO; ZHOU; HUANG, 2009). Without this, any attempt to classify the data using a small number of groups would be meaningless. K-means Cluster Analysis, applied to the 13 first PCs was able to classify all *solo* executions into 6 groups, corresponding to the 6 players with 100% accuracy. This may indicate that each performer have a consistent way of moving, which appears to be distinct from the others. This suggests the existence of individual "gesture signatures", corroborating the findings of previous studies. Figure 5.4 shows the partition of the first two principal components into 6 players with 100% accuracy.

Gestural coupling

Having demonstrated that the kinematic data contains sufficient recurrence to identify individual "gestural signatures" of musicians, we

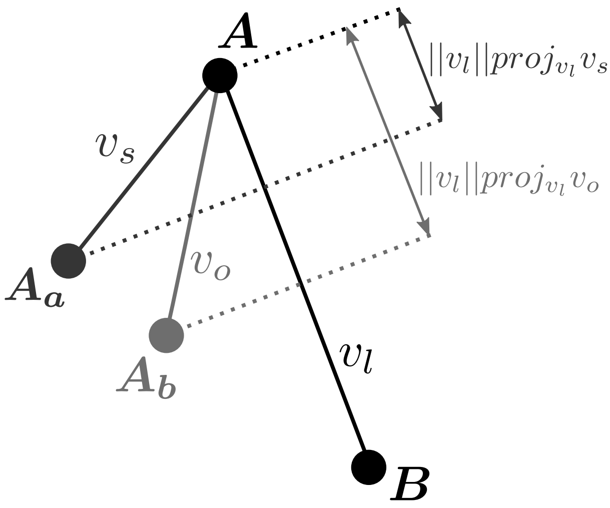

Figure 5.5: Geometric illustration of the vector projection procedure. The points in the *velocity profile* space corresponding to the solo performances of the first and the second clarinetists are indicated by $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. The performance of the first clarinetist following himself is represented by $\boldsymbol{A_a}$, while $\boldsymbol{A_b}$ represents the performance following the other clarinetist.

tried to verify if musicians would change their "gestural signatures" by influence of different interpretative conditions imposed by the musical conception of a leader. The adaptation of the followers' gestures to those of the leaders was evaluated by projecting their *velocity profiles*, while playing as second clarinetist, onto the dimension that separate the leaders apart.

This was done as follows. First, we represent the leader performances of two clarinetists in the space of velocity profiles (points $\boldsymbol{A}$ and $\boldsymbol{B}$ in Figure 5.5). The vector that connects the point $\boldsymbol{A}$ to the point $\boldsymbol{B}$ is denoted $\boldsymbol{v_l}$. We then consider the performance of the first clarinetist when playing with her- or himself (point $\boldsymbol{A_a}$) and when playing with the second clarinetist (point $\boldsymbol{A_b}$). The amount of change in the kinematic pattern is evaluated by computing the projections of the vector $\boldsymbol{A} \longleftrightarrow \boldsymbol{A_a}$ ($\boldsymbol{v}_s$, the "self" condition) and the vector $\boldsymbol{A} \longleftrightarrow \boldsymbol{A_b}$ ($\boldsymbol{v}_o$, the "other" condition) onto the vector $\boldsymbol{v_l}$. These are the vectors represented as $proj_{v_l}\boldsymbol{v}_s$ and $proj_{v_l}\boldsymbol{v}_o$, respectively, according to equation (5.1).

$$proj_{v_l}\boldsymbol{v}_s = \frac{\boldsymbol{v}_s \cdot \boldsymbol{v}_l}{||\boldsymbol{v}_l||} \qquad proj_{v_l}\boldsymbol{v}_o = \frac{\boldsymbol{v}_o \cdot \boldsymbol{v}_l}{||\boldsymbol{v}_l||} \tag{5.1}$$

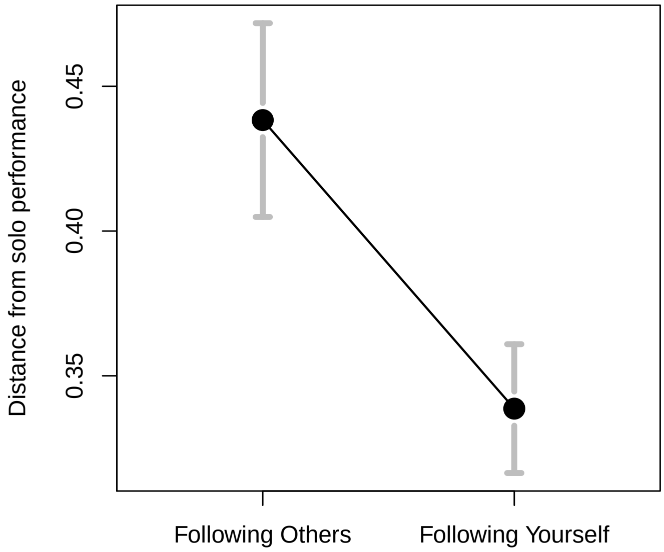

Figure 5.6: Distance from solo gestural signature in Self and Other performance conditions. Mean values are shown with dots and the standard error bars are shown in gray.

The length of the projected vectors were adopted as a metric to differentiate executions where musicians follow themselves from those where they follow others. Small values of $proj_{v_l}v_s$ indicate that musicians maintain their gestural signature when accompany her/his own recordings, while increased values of $proj_{v_l}v_o$ indicate that they would abandon their own gestural signature to adjust to that of the other.

A two-tailed t-test with Welch's correction was applied to the *self* and *other* conditions. In total, 240 vector projections were considered, they correspond to four takes of the two projections in each of the 30 pairs of clarinetists, ordered without repetition ($n(n-1)$). Results indicated significant mean difference between $proj_{v_l}v_s$ ($M = 0.338, SD = 0.118$) and $proj_{v_l}v_o$ ($M = 0.438, SD = 0.173$); $t(184.3) = 4.91$, $p < .0001$, suggesting that musicians tended to maintain their original gestural profile while following their own executions, but shifted towards others' gestures profiles when following them (Figure 5.6).

Results of a one-way ANOVA performed on the subgroup *self* suggests a significant decrease of the distance along subsequent takes, indicating an adaptation towards the original "solo" gestural signature, $F(3,108) = 5.054, p < .01$, as shown in Figure 5.6 left panel, while no significant differences related to takes, $F(3,102) = 0.097, p > .96$ was observed for the subgroup *other* (right panel). This might additionally argue towards the existence of individual "gestural signatures" in musical performances.

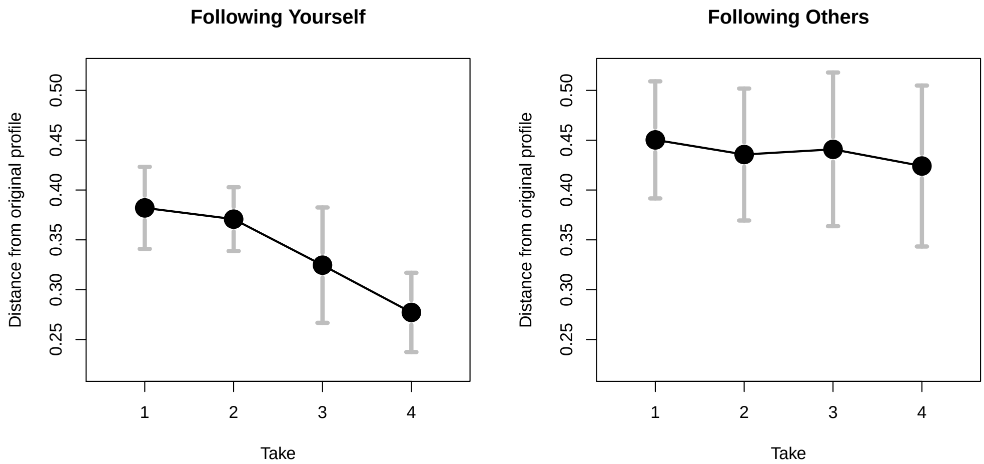

Figure 5.7: Distance from solo gestural signature in Self (left panel) and Other (right panel) performance conditions across takes. Mean values are shown with dots and the standard error bars are shown in gray.

5.2.2 *Discussion*

This investigation aimed at determining whether body movement of musicians contains information related to the interpretative intention during a performance, by verifying the occurrence of gestural coupling in clarinet duos. Our experiments enabled us to investigate the influence of different interpretive situations on individual patterns of body movement. The procedures used aimed at identifying gestural pattern recurrence in solo performances, which we considered as individual "gesture signatures", and to verify if those "signatures" would be affected when the performer followed recordings of other musicians.

The previously discussed *self-other* effect was adopted as a framework for our experiment design, as it was mentioned in earlier studies on synchronization of music ensembles, (KELLER, 2001; KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012). The most accepted hypothesis for explaining the *self-other* effect suggests that the coordination between musicians is achieved by each interpreter internally simulating the actions of other members of the group, initially relying on how they would perform the music themselves. Hence, when following their own recorded performances, they would recognize the musical actions of the first clarinetist as self-produced and would more easily adapt to them.

In the present experiment, we tested the existence of gesture coupling between two clarinetists playing duets in unison, without any visual contact between them. First, we were able to recognize musicians in solo performances with 100% accuracy. Next, we were able to observe that clarinetists, when following themselves, tended to retain their "original" gestural signature, as recorded in solo performances. Yet, when following others, it was observed that they tended to deviate from it. Moreover, our results indicates a tendency of adaptation between the gestural signature of the follower to that of the leader, even without visual contact with the partner.

The observed gestural variations observed in the different performance conditions could indicate an involuntary attempt of musicians to anticipate the interpretive intentions of the leader, only by hearing her/his manipulations of acoustic parameters. Therefore, these results are favorable to the hypothesis that ancillary gestures contain some sort of musical significance, corroborating what has been demonstrated by several studies, such as the experiments carried by Caramiaux, Bevilacqua and Schnell (2010), which showed that parameters extracted from the gestures of listeners, such as position, velocity, normal and tangential accelerations, curvature, radius and torsion, were correlated with acoustic parameters, such as loudness and sharpness.

5.2.3 *Conclusion*

We propose a multimodal analysis framework intended to access the interactions between gesture and music in ensemble performances. Human communication is not limited to the use of only one form of sensory information. Just as in verbal communication, music makes use of various mechanisms not directly associated with the production of sound as a means of communicating the desired interpretation, such as the movements of the body, or the facial and gesture expressions of a conductor. We believe that the gestural recurrence observed between performances of the same musician is strongly connected to the planning of the interpretation to be realized. On the other hand, by considering that musicians tend to bend their gestures towards the gesture of the leader, we might hypothesize that such gestural adjustments would reflect her/his ability of musically "fitting" to performances of others musicians. As musicians in an orchestra commonly get used to the gestures of a conductor, it might be possible to assume that musicians are able to learn how to read the movement of their partners. Further multimodal investigations of musical ensemble performances may facilitate the comprehension of the creative process underlying musical interpretation, which could shed light on questions such as: why some musicians feel more comfortable playing with a particular partner, and would gestural information contribute to this? We think that the framework proposed in this investigation would contribute to answer these questions.

6

Conclusion and Future Works

In this work we tried to address the issue of adaptation between leaders and followers in clarinet duets, by focusing on acoustical and gestural responses yielded by musicians during solo and ensemble performances. Firstly, we presented an overview about ensemble performance and gesture-related research literature, focusing on synchronization and coordination in musical ensemble, discussing their importance to the field. We then presented the experimental procedures used in the work, discussing specifically the onset detection strategies available and how they can influence the results of time related musical research. We also discussed what this field of research could gain from a move towards more organic experiment setups that are more relatable to real performances scenarios.

We chose to approach the topic of coordination in ensemble performance in multiple studies, each one focused on different aspects of this interaction. In Chapter 4 we focused on the rhythmic aspects of ensemble interactions, specifically, in Section 4.1 we tried to move forward in investigating in detail which factors influence leader/follower interactions and which factors would make a musician an ideal co-performer. We were able to verify the occurrence of the *self-other* effect in different musical conditions. The most accepted hypothesis to explain the *self-other* effect suggests that coordination between musicians may be achieved by each interpreter internally simulating the actions of the other members of the group, relying initially on how they would perform the excerpt (KELLER, 2001; KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012). Therefore, when following themselves, they would recognize the musical actions of the first clarinetist as self-produced, which would enable them to a better adaptation. The *self-other* effect observed in this investigation, showed results consistent with those reported by Keller, Knoblich and Repp (2007), corroborating the idea that musicians tend to follow their own executions more efficiently. We were also able to observe evidences that 1) some participants exhibit better accompanying skills than others; 2) some participants induce more asynchrony to their co-performers, i.e. they are harder to follow (induced asynchrony effect); and 3) some participants coordinate better with certain co-performers.

The section discussed the synchronization in clarinet duos on different performance scenarios. We tried to observe how musicians adapt their performances to different partners. We observed that each musician assuming the role of follower responds differently to the task, suggesting that different musicians have different ability levels as follower. In parallel, a similar result was observed for the musicians in the role of leader. Some participants induced greater asynchrony in the performance of their colleagues, while others induced less. We also discussed which elements may influence this behavior. When the musicians played following their own performances it was expected that they would achieve a greater synchrony, as observed in the results. But even in those cases there is still an attempt to adjust to the recording, which occurs note by note. This suggests that the familiarity with the recorded performance facilitates the understanding of the musical information being transmitted and increases the predictability of the musical ideas of the leader. We also discussed the concept of a musical signature. We hypothesized the existence of a musical signature for each performer and that musicians with similar musical signatures would have a better chance to adjust to each other. In this context, the best adjustment case would be when the musician plays following his own previously recorded performance.

In the following section (4.2) we further explored the concept of musical signature, discussing how they would emerge and influence on the adjustment of musical groups. We investigated repetitive rhythmic patterns with various occurrences in the excerpt and tried to identify rhythmic signatures that could be occurring. The results of a MANOVA for pairs of clarinetists suggests the existence of individual rhythmic signatures, which is surprising for a short rhythmic pattern of only three notes. This result also highlighted the different values of consistency in the repetition of rhythmic patterns for each musician. Afterwards, we tried to verify if there was any relationship between these rhythmic signatures, their consistency and the observed asynchrony induced by leaders. We observed that musicians that induce greater asynchrony are less consistent on the realization of repeated rhythmic patterns which, in the light of the Common-Coding Theory (HOMMEL et al., 2001), would indicate that they are less predictable for their co-performers. The hypothesis raised here was that **leaders that play more consistently induce less asynchrony** on their followers. This notion seems to be a common sense among performing musicians and was suggested in studies like Repp (1995), although there is still no sufficient data to support this assumption.

In Chapter 5 we addressed the gestural component of ensemble musical performance to investigate how gestural information would contribute to the synchronization in musical groups. In section 5.1, we tried to extend the concept of "musical signature" to "gestural signatures". This reasoning was driven by the high gestural recurrence observed during the performances. Further, we investigated if the interaction with different musicians would disrupt this "gestural signatures" when following others. We observed that musicians presented distinctive, highly consistent, patterns of body movement while performing, which fits well with the concept of gestural signature proposed in this study. Moreover, we observed that musicians gestural signatures are disrupted when they play in duets, suggesting that some gesture adaptation may be taking place.

When comparing performances of six different musicians, regions of inter-performer commonality, i.e regions where all musicians performed similar gestures, were detected. Those regions tended to occur in positions of great musical significance, which strongly suggests a connection of such behavior to the musical structure. Furthermore, inside those regions of commonality, we were still able to detect individual characteristics. Duet performances were investigated in search for disturbances of gestural signatures of the followers. Results indicated that gesture signatures were less disturbed when musicians followed their own recordings, as compared to the situation when they followed recordings made by others, a similar effect already observed on acoustic asynchrony.

And, finally, in section 5.2 we further explored the hypothesis of the gestural signatures by applying an alternative approach with a different musical excerpt, and tested if the disruption observed in followers signatures were related to the signature of the leaders. The results of kinematic data analysis indicate that when participants followed themselves they tended to retained their original gestural profile, as recorded in solo executions. However, when they followed other clarinetists, they tended to deviate from their original profile. Moreover, it was observed that the gesture patterns of followers tended to slightly adapt to those of leader clarinetists. Finally, since there were no visual interaction between the participants during recordings, the observed behavior could suggest a connection between variations in the gestural patterns and the manipulation of musical parameters carried out by leader clarinetists. In other words, when clarinetists followed the leaders they tried to adapt their performance to the intentions of the other, consequently, it is possible that the observed changes in gestural patterns could be a reflection of this attempt. These results are favorable to the hypothesis of coupling between gesture and music. Several studies have demonstrated evidence of this relationship, such as Glowinski et al. (2013), Palmer et al. (2009b), Amelynck et al. (2014), Keller and Appel (2010), Dahl and Friberg (2007). We believe that the gestural recurrence observed between performances of the same musician is strongly connected to the planning of the interpretation to be realized. On the other hand, by considering that musicians tend to bend their gestures towards the gesture of the leader, we might hypothesize that such gestural adjustments would reflect her/his ability of musically "fitting" to performances of others musicians. As musicians in an orchestra commonly get used to the gestures of a conductor, it might be possible to assume that musicians are able to learn how to read the movement of their partners. Further multimodal investigations of musical ensemble performances may facilitate the comprehension of the creative process underlying musical interpretation, which could shed light on questions such as: why some musicians feel more comfortable playing with a particular partner, and would gestural information contribute to this?

6.1 GENERAL DISCUSSION

Informally, we may say that the main question driving this study is "who is your ideal co-performer, and why"? A fundamental concept for this approach is that of the signature, which we applied to both acoustic and gestural domains. Different authors make reference to what we call "musical signatures" of interpreters. Most of them indicate that those signatures are linked to the interpretative intentions of musicians, and are usually observed in solo performances (REPP, 1992; REPP, 1995; REPP, 1996; GOODMAN, 2002; KOREN; GINGRAS, 2014). Similarly, we have observed the manifestation of these signatures in all performers analyzed in this work, which supports previous findings in the literature. The results presented in Section 4.2 have demonstrated evidence for consistent, individual musical signatures that are expressed not only in higher levels of the musical performance (phrase organization, articulations, etc.) but also in small micro-timing variations inherent to each performer.

Other studies supports that movements performed by musicians during musical performances are consistent and, to a certain extent, reflect some relation with the musical characteristics of the performed music (JENSENIUS et al., 2010; CARAMIAUX; WANDERLEY; BEVILACQUA, 2012; TEIXEIRA et al., 2015). In previous works (MOTA, 2012; MOTA; LOUREIRO; LABOISSIèRE, 2013), we verified that these movements, besides being consistent, exhibit unique characteristics of each performer, allowing them to be identifiable and, therefore, to function as a signature for each musician. The idea of gesture signatures is the focus of several studies that discuss its relevance and application for multiple purposes, such as personal identification systems (FARELLA et al., 2006), human computer interaction (BEVILACQUA et al., 2010) and the recognition of humans and their activities (LOULA et al., 2005; CHELLAPPA; ROY-CHOWDHURY; ZHOU, 2005). Likewise, many studies in musical performance observed the recurrence of musicians gestural patterns (NUSSECK; WANDERLEY, 2009; DESMET et al., 2012; TEIXEIRA et al., 2015) and discussed its use for musical interfaces (WANDERLEY et al., 2005; CARAMIAUX; WANDERLEY; BEVILACQUA, 2012).

When playing in groups it is assumed that the musical signature of an interpreter is altered in order to adapt to the expressive choices of the other members of the group, which could be verified in different studies (see: Shaffer (1984), Goodman (2002) and Keller (2014)). However, in the case of gestural signatures this process seems more complex, because they can reflect musical characteristics of the performed music and be replete with gestural idiosyncrasies of each interpreter at the same time. If we consider the gesture-music relationship as a multimodal amalgam in the light of theories such as embodied cognition (LEMAN, 2007), it is expected that during ensemble performances a personal signature would present changes reflecting the musical choices of the other members of the group, because of their mutual musical or gestural influence. On the other hand, if we consider the movement of the musicians as a reflection of the aforementioned gestural idiosyncrasies of each individual, we would expect that these signatures would be unchanged despite the different performance condition. Nonetheless, what we have observed in this study is somewhat of a mixture between the two hypotheses. Not only, we observed that certain gestural characteristics, specific to each performer, were maintained throughout ensemble performances but, also, some changes reflecting musical adaptation could be verified, similar to what happened to musical signatures.

Indeed, what we observed is that musicians do not entirely give up their gestural signature, even when following others. Yet, the results indicate that this gestural signature is disturbed when musicians have to waive their musical signature, supporting the connection between the two modalities. One strong hypothesis derived from this work is that the so called ancillary gestures could have a function of firstly, serving as a reference frame for the acting musician, and secondly transmitting this reference frame for co-performers.

In conclusion, we believe that applications of this research may emerge for educational environments where computational visualization tools can help raise awareness of the movements being performed and their effect on the musical coupling in ensemble performance. Other possible applications would be new interfaces for digital musical instruments, or intelligent accompanying systems, able to "follow" gestural and acoustic cues using communication strategies similar to those employed by real musicians. It is also possible to elaborate performances with synthesized movements using sampling from elements extracted from real musicians (e.g. average movements of several musicians). Will those performances be identified as real? Do they influence the coupling between musicians?

6.2 FUTURE WORKS

We are currently in the process of investigating whether changes in the kinematic representation of musicians in the role of *leader* may influence the perception of musical parameters by co-performers. In ensemble music performance, synchronization can be achieved by means of acoustical or visual streams, but a number of studies, such as Repp (2005), Repp and Su (2013), Wöllner and Cañal-Bruland (2010) shows that the combination of these sensory modalities can improve the overall synchronization of an ensemble. We believe that the coherence between these modalities should play key role in this process. A study conducted by Nusseck and Wanderley (2009) indicated that changing the kinematic properties (amplitude of the movements) of a musician movements can influence the perceptual impressions of her/his performance. In the study, the authors used kinematic displays, a stick figure representation of musicians' body. They ask subjects to rate specific music-related dimensions of the performances (perceived tension, intensity, fluency and professionalism). Results showed that participants judged as more intense the interpretations with the range of motion digitally enhanced, even without any changes in the audio streams. From that perspective, we aim at investigating the contribution of ancillary body movements to the synchronization of musical ensembles.

The question that we ask is: to what extent gesture information influence the synchronization of musical tasks? To answer this, we propose to test if changes in the kinematic representation of a musician can affect the overall synchronization of their peers musicians. We propose to test if the absence of gestural information and misleading gestural information have an effect in the synchronization of musically experienced subjects. We are focusing on testing if the incongruence of modalities (mixing gesture from one musician and the audio from another) would have any effect in the synchronization results of a third player. We present the visual stimulus combined to a variable mixture of recorded audio and a masker signal. We regulate the amount of musical information the listener will have access by controlling the Signal to Noise Ratio (SNR) in the audio streams, ranging from highly noise to no noise. Our hypothesis are: 1) in incongruent conditions it would be harder to follow; 2) at a given SNR the effect of congruence would cease to exist, rendering similar asynchrony values for the curves of congruent and incongruent conditions. We believe that coherence between these modalities should play a central role in this process. Consequently, disrupting this coherence we expect to see a decrease in the overall synchronization registered. Furthermore, by tricking participants' cognitive processes we expect to see if participants choose to waive one modality in favor of the other.

References

AMELYNCK, D. et al. Expressive body movement responses to music are coherent, consistent, and low dimensional. *Cybernetics, IEEE Transactions on*, v. 44, n. 12, p. 2288–2301, Dec 2014. ISSN 2168-2267.

ATIK, Y. The conductor and the orchestra: Interactive aspects of the leadership process. *Leadership & Organization Development Journal*, v. 15, n. 1, p. 22–28, 1994. Available at: <http://dx.doi.org/10.1108-/0e1437739410050123>.

BATES, D. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, n. 1, p. 1–48, 2015.

BELLO, J. et al. A tutorial on onset detection in music signals. *Speech and Audio Processing, IEEE Transactions on*, v. 13, n. 5, p. 1035–1047, 2005.

BEVILACQUA, F. et al. Continuous realtime gesture following and recognition. In: KOPP, S.; WACHSMUTH, I. (Ed.). *Gesture in Embodied Communication and Human-Computer Interaction*. Springer Berlin Heidelberg, 2010, (Lecture Notes in Computer Science, v. 5934). p. 73–84. ISBN 978-3-642-12552-2. Available at: <http://dx.doi.org/10.1007/978-3-642-12553-9_7>.

BISHOP, L.; GOEBL, W. Context-specific effects of musical expertise on audiovisual integration. *Frontiers in Psychology*, v. 5, n. 1123, 2014. ISSN 1664-1078. Available at: <http://www.frontiersin.org-/cognitive_science/10.3389/fpsyg.2014.01123/abstract>.

CADOZ, C.; WANDERLEY, M. et al. Gesture-music. *Trends in Gestural Control of Music*, Paris: Ircam, p. 71–94, 2000.

CARAMIAUX, B.; BEVILACQUA, F.; SCHNELL, N. Towards a gesture-sound cross-modal analysis. In: KOPP, S.; WACHSMUTH, I. (Ed.). *Gesture in Embodied Communication and Human-Computer Interaction*. Springer Berlin Heidelberg, 2010, (Lecture Notes in Computer

Science, v. 5934). p. 158–170. ISBN 978-3-642-12552-2. Available at: <http://dx.doi.org/10.1007/978-3-642-12553-9_14>.

CARAMIAUX, B.; WANDERLEY, M.; BEVILACQUA, F. Segmenting and parsing instrumentalists' gestures. *Journal of New Music Research*, Taylor & Francis, v. 41, n. 1, p. 13–29, 2012.

CHELLAPPA, R.; ROY-CHOWDHURY, A. K.; ZHOU, S. K. *Recognition of Humans and Their Activities Using Video*. [S.l.]: Morgan Claypool Publishers, 2005. 1-179 p. ISBN 1598290061.

CLYNES, M. Microstructural musical linguistics: composers' pulses are liked most by the best musicians. *Cognition*, Elsevier, v. 55, n. 3, p. 269–310, 1995.

DAHL, S.; FRIBERG, A. Visual perception of expressiveness in musicians' body movements. *Music Perception*, JSTOR, v. 24, n. 5, p. 433–454, 2007.

DELALANDE, F. La gestique de gould. *Glenn Gould Pluriel*, p. 85–111, 1988.

DESMET, F. et al. Assessing a clarinet player's performer gestures in relation to locally intended musical targets. *Journal of New Music Research*, v. 41, n. 1, p. 31–48, 2012.

DIXON, S. Onset detection revisited. In: CITESEER. *Proceedings of the 9th International Conference on Digital Audio Effects*. [S.l.], 2006. v. 120, p. 133–137.

FAIRHURST, M. T.; JANATA, P.; KELLER, P. E. Leading the follower: an fmri investigation of dynamic cooperativity and leader–follower strategies in synchronization with an adaptive virtual partner. *Neuroimage*, Elsevier, v. 84, p. 688–697, 2014.

FARELLA, E. et al. Gesture signature for ambient intelligence applications: A feasibility study. In: FISHKIN, K. et al. (Ed.). *Pervasive Computing*. [S.l.]: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 3968). p. 288–304. ISBN 978-3-540-33894-9.

GABRIELSSON, A. Music performance research at the millennium. *Psychology of music*, Sage Publications, v. 31, n. 3, p. 221–272, 2003.

GALLESE, V. et al. Action recognition in the premotor cortex. *Brain*, Oxford University Press, v. 119, n. 2, p. 593–609, 1996. ISSN 0006-8950. Available at: <http://brain.oxfordjournals.org/content/119/2/593>.

GINGRAS, B. et al. Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, Pion Ltd, v. 40, n. 10, p. 1206–1220, 2011. Available at: <http://www.perceptionweb.com-/abstract.cgi?id=p6891>.

GLOWINSKI, D. et al. Expressive non-verbal interaction in a string quartet: an analysis through head movements. *Journal on Multimodal User Interfaces*, Springer Berlin Heidelberg, p. 1–14, 2014. ISSN 1783-7677. Available at: <http://dx.doi.org/10.1007/s12193-014-0154-3>.

GLOWINSKI, D. et al. The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves. *Frontiers in Psychology*, v. 4, p. 841, 2013. ISSN 1664-1078. Available at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00841>.

GOEBL, W.; PALMER, C. Synchronization of timing and motion among performing musicians. *Music Perception*, JSTOR, v. 26, n. 5, p. 427–438, 2009.

GOODMAN, E. Musical performance: A guide to understanding. In: ____. [S.l.]: Cambridge University Press, 2002. chap. Ensemble performance, p. 153–167.

HOMMEL, B. et al. The theory of event coding (tec): A framework for perception and action planning. *Behavioral and Brain Sciences*, v. 24, p. 849–878, 10 2001. ISSN 1469-1825. Available at: <http://journals-.cambridge.org/article_S0140525X01000103>.

HURON, D. Note-onset asynchrony in js bach's two-part inventions. *Music Perception: An Interdisciplinary Journal*, University of California Press Journals, v. 10, n. 4, p. 435–443, 1993.

JEANNEROD, M. Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, Elsevier, v. 14, n. 1, p. S103–S109, 2001.

JENSENIUS, A. R. et al. Musical gestures: concepts and methods in research. In: ____. [S.l.]: Routledge, 2010. (Musical gestures : sound, movement, and meaning), p. 12–35. ISBN 9780415998871.

KELLER, P. Attentional resource allocation in musical ensemble performance. *Psychology of Music*, SAGE Publications, v. 29, n. 1, p. 20–38, 2001.

KELLER, P.; APPEL, M. Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, JSTOR, v. 28, n. 1, p. 27–46, 2010.

KELLER, P.; KNOBLICH, G.; REPP, B. Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and cognition*, Elsevier, v. 16, n. 1, p. 102–111, 2007.

KELLER, P. E. Joint action in music performance. In: MORGANTI, F.; CARASSA, A.; G., R. (Ed.). *Emerging Communication: Studies in New Technologies and Practices in Communication*. Amsterdam: IOS press, 2008. v. 10, p. 205–221.

KELLER, P. E. Ensemble performance: interpersonal alignment of musical expression. In: FABIAN, D.; TIMMERS, R.; SCHUBERT, E. (Ed.). *Expressiveness in music performance: Empirical approaches across styles and cultures*. [S.l.]: Oxford University Press New York, NY, 2014. p. 260–282.

KNOBLICH, G.; FLACH, R. Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, SAGE Publications, v. 12, n. 6, p. 467–472, 2001.

KNOBLICH, G. et al. Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *The Quarterly Journal of Experimental Psychology: Section A*, Taylor & Francis, v. 55, n. 3, p. 1027–1046, 2002.

KOREN, R.; GINGRAS, B. Perceiving individuality in harpsichord performance. *Frontiers in psychology*, Frontiers Media SA, v. 5, 2014.

KUZNETSOVA, A.; Bruun Brockhoff, P.; Haubo Bojesen Christensen, R. *lmerTest: Tests in Linear Mixed Effects Models*. [S.l.], 2016. R package version 2.0-30. Available at: <https://CRAN.R-project.org/package=lmerTest>.

LEMAN, M. *Embodied music cognition and mediation technology*. Cambridge, Massachusetts: Mit Press, 2007.

LEMAN, M.; NAVEDA, L. Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in samba and charleston. *Music Perception: An Interdisciplinary Journal*, University of California Press, v. 28, n. 1, p. 71–91, 2010. ISSN 07307829, 15338312. Available at: <http://www.jstor.org/stable/10.1525/mp.2010.28.1.71>.

LEVEAU, P.; DAUDET, L. Methodology and tools for the evaluation of automatic onset detection algorithms in music. In: *In Proc. Int. Symp. on Music Information Retrieval*. [S.l.: s.n.], 2004. p. 72–75.

LOULA, F. et al. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, American Psychological Association, v. 31, n. 1, p. 210, 2005.

LOUREIRO, M. et al. Acoplamento interpretativo e sincronização em duos de clarinetas. *OPUS*, v. 18, n. 2, p. 181–206, 2012. ISSN 1517-7017. Available at: <http://www.anppom.com.br/opus/data/issues-/archive/18.2/files/OPUS_18_2_Loureiro_et_al.pdf>.

LOUREIRO, M. A.; CAMPOLINA, T.; MOTA, D. A. Expan: a tool for musical expressiveness analysis. In: *Proceedings of the 2nd International Conference of Systematic Musicology (Sys Mus2009)*. Ghent, Belgium: Ugent, 2009. p. 24–27. Available at: <http://www.ipem.ugent.be-/sysmus09/files/Sysmus09_webproceedings.pdf>.

LOUREIRO, M. A. et al. Evidence for self-other effects and structure-related attentional mechanisms in the primo/secondo performances of clarinetists. In: *Proceedings of the 16th International Conference on Perception and Action (ICPA 2011)*. [S.l.: s.n.], 2011.

LOUREIRO, M. A. et al. Padrões de sincronização temporal em duos de clarinetas: influencia do acompanhante e da estrutura musical. In: *Anais do XVIII Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música*. João Pessoa, Brasil: Universidade Federal do Paraíba, 2012. p. 581–585. Available at: <http://www.anppom.com.br/anais-/anaiscongresso_anppom_2012/VIII_MusicaeTecnologiaeSonologia-.pdf>.

LOUREIRO, M. A. et al. Content analysis of note transitions in music performance. In: GOUYON, F.; BARBOSA, A.; SERRA, X. (Ed.). *Proceedings of the 6th Sound and Music Computing Conference Sound and Music Computing Conference*. Porto: INESC, 2009. p. 355–359. Available at: <http://smc2009.smcnetwork.org/programme/pdfs/291.pdf>.

MAHALANOBIS, P. C. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, v. 2, p. 49–55, 1936.

MARCHINI, M. et al. The sense of ensemble: a machine learning approach to expressive performance modelling in string quartets. *Journal of New Music Research*, v. 43, n. 3, p. 303–317, 2014. Available at: <http://dx.doi.org/10.1080/09298215.2014.922999>.

MEISTER, I. et al. Playing piano in the mind—an fmri study on music imagery and performance in pianists. *Cognitive Brain Research*, v. 19, n. 3, p. 219 – 228, 2004. ISSN 0926-6410. Available at: <http://www-.sciencedirect.com/science/article/pii/S0926641004000023>.

MOTA, D.; LOUREIRO, M.; LABOISSIèRE, R. Analysis of synchronization patterns in clarinet duos using acoustic and kinematic parameters. In: SHIFRES, F. et al. (Ed.). *Actas de ECCoM. Vol. 1 No1, "Nuestro Cuerpo en Nuestra Música. 11o ECCoM"*. Buenos Aires: SACCoM, 2013. v. 1, n. 1, p. 199–206. ISSN 2346-8874. Available at: <http://www.saccom.org.ar-/actas_eccom/vol1-1_contenido/mota_loureiro_y_laboissiere.pdf>.

MOTA, D. A. *Análise dos padrões de sincronização em duos de clarineta a partir de parâmetros acústicos e cinemáticos*. Master's Thesis (Master's Thesis) — Universidade Federal de Minas Gerais, 2012.

NETTHEIM, N.; BECKING, G. How musical rhythm reveals human attitudes: Gustav becking's theory. *International Review of the Aesthetics and Sociology of Music*, JSTOR, p. 101–122, 1996.

NOVEMBRE, G.; KELLER, P. E. A conceptual review on action-perception coupling in the musicians' brain: what is it good for? *Frontiers in Human Neuroscience*, v. 8, p. 603, 2014. ISSN 1662-5161. Available at: <http://journal.frontiersin.org/article/10.3389/fnhum-.2014.00603>.

NUSSECK, M.; WANDERLEY, M. M. Music and motion—how music-related ancillary body movements contribute to the experience of music. *Music Perception: An Interdisciplinary Journal*, University of California Press, v. 26, n. 4, p. 335–353, 2009. ISSN 07307829. Available at: <http://www.jstor.org/stable/10.1525/mp.2009.26.4.335>.

PALMER, C. Music performance. *Annual Review of Psychology*, v. 48, n. 1, p. 115–138, 1997. PMID: 9046557. Available at: <http://www-.annualreviews.org/doi/abs/10.1146/annurev.psych.48.1.115>.

PALMER, C. et al. Synchronization of motion and timing in clarinet performance. In: *International Symposium on Performance Science*. [S.l.: s.n.], 2009. p. 1–6.

PALMER, C. et al. Movement-related feedback and temporal accuracy in clarinet performance. *Music Perception: An Interdisciplinary Journal*, University of California Press, v. 26, n. 5, p. 439–449, 2009. ISSN 07307829, 15338312. Available at: <http://www.jstor.org/stable/10.1525/mp-.2009.26.5.439>.

PALMER, C. et al. Temporal coordination in vocal duet performances of musical rounds. In: *Proceedings of the Sound and Music Computing (SMC) Conference*. [S.l.: s.n.], 2013.

QIAO, Z.; ZHOU, L.; HUANG, J. Z. Sparse linear discriminant analysis with applications to high dimensional low sample size data. *IAENG International Journal of Applied Mathematics*, v. 39, n. 1, p. 06, 2009.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Available at: <https://www.R-project.org/>.

RASAMIMANANA, N. Towards a conceptual framework for exploring and modelling expressive musical gestures. *Journal of New Music Research*, v. 41, n. 1, p. 3–12, 2012.

RASCH, R. A. The perception of simultaneous notes such as in polyphonic music. *Acta Acustica united with Acustica*, S. Hirzel Verlag, v. 40, n. 1, p. 21–33, 1978.

RASCH, R. A. Synchronization in performed ensemble music. *Acta Acustica united with Acustica*, v. 43, n. 2, p. 121–131, 1979. Available at: <http://www.ingentaconnect.com/content/dav/aaua/1979-/00000043/00000002/art00005>.

REPP, B. Expressive timing in schumann's "träumerei:"an analysis of performances by graduate student pianists. *The Journal of the Acoustical Society of America*, v. 98, p. 2413–2427, 1995.

REPP, B. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin Review*, Springer-Verlag, v. 12, n. 6, p. 969–992, 2005. ISSN 1069-9384. Available at: <http://dx.doi.org/10-.3758/BF03206433>.

REPP, B. H. Diversity and commonality in music performance: An analysis of timing microstructure in schumann's "tr[a-umlaut]umerei". *The Journal of the Acoustical Society of America*, ASA, v. 92, n. 5, p. 2546–2568, 1992. Available at: <http://link.aip.org/link/?JAS/92/2546/1>.

REPP, B. H. Music as motion: A synopsis of alexander truslit's (1938) gestaltung und bewegung in der musik. *Psychology of Music*, Sage Publications, v. 21, n. 1, p. 48–72, 1993.

REPP, B. H. Musical motion: Some historical and contemporary perspectives. In: KGL. MUSIKALISKA AKADEMIN STOCKHOLM. *Proceedings of the Stockholm Music Acoustics Conference (SMAC)*. [S.l.], 1993. p. 128–135.

REPP, B. H. Pedal timing and tempo in expressive piano performance: A preliminary investigation. *Psychology of Music*, v. 24, n. 2, p. 199–221, 1996. Available at: <http://pom.sagepub.com/content/24/2/199.abstract>.

REPP, B. H. The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, JSTOR, p. 419–444, 1997.

REPP, B. H.; SU, Y.-H. Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, v. 20, n. 3, p. 403–452, 2013. Available at: <http://dx.doi.org/10.3758/s13423-012-0371-2>.

REPP, B. H.; WINDSOR, W. L.; DESAIN, P. Effects of tempo on the timing of simple musical rhythms. *Music Perception: An Interdisciplinary Journal*, University of California Press Journals, v. 19, n. 4, p. 565–593, 2002. ISSN 0730-7829. Available at: <http://mp.ucpress.edu/content/19/4/565>.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 1986.

SEBANZ, N.; BEKKERING, H.; KNOBLICH, G. Joint action: bodies and minds moving together. *Trends in cognitive sciences*, Elsevier, v. 10, n. 2, p. 70–76, 2006.

SHAFFER, L. H. Timing in solo and duet piano performances. *The Quarterly Journal of Experimental Psychology*, Taylor & Francis, v. 36, n. 4, p. 577–595, 1984.

SPITZER, J.; ZASLAW, N. *The birth of the orchestra: history of an institution, 1650-1815*. [S.l.]: Oxford University Press, 2004.

TEIXEIRA, E. C. et al. Motion analysis of clarinet performers. *Journal of New Music Research*, Taylor & Francis, v. 44, n. 2, p. 97–111, 2015.

TIMMERS, R. et al. Synchronization and leadership in string quartet performance: a case study of auditory and visual cues. *Frontiers in psychology*, Frontiers Media SA, v. 5, 2014.

TODD, M.; NEIL, P. The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, v. 91, n. 6, p. 3540–3550, 1992.

TODD, N. A computational model of rubato. *Contemporary Music Review*, Taylor & Francis, v. 3, n. 1, p. 69–88, 1989.

TODD, N. Towards a cognitive theory of expression: The performance and perception of rubato. *Contemporary Music Review*, Taylor & Francis, v. 4, n. 1, p. 405–416, 1989.

TODD, N. Motion in music: A neurobiological perspective. *Music Perception*, JSTOR, p. 115–126, 1999.

TSAY, C.-J. The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, v. 124, n. 1, p. 24 – 33, 2014. ISSN 0749-5978. Available at: <http://www.sciencedirect-.com/science/article/pii/S0749597813000861>.

VOS, J.; RASCH, R. The perceptual onset of musical tones. In: ____. *Music, Mind, and Brain: The Neuropsychology of Music*. Boston, MA: Springer US, 1982. p. 299–319. ISBN 978-1-4684-8917-0. Available at: <http://dx.doi.org/10.1007/978-1-4684-8917-0_16>.

WANDERLEY, M. Non-obvious performer gestures in instrumental music. In: BRAFFORT, A. et al. (Ed.). *Gesture-Based Communication in Human-Computer Interaction*. [S.l.]: Springer Berlin Heidelberg, 1999, (Lecture Notes in Computer Science, v. 1739). p. 37–48. ISBN 978-3-540-66935-7.

WANDERLEY, M. Quantitative analysis of non-obvious performer gestures. In: WACHSMUTH, I.; SOWA, T. (Ed.). *Gesture and Sign Language in Human-Computer Interaction*. Springer Berlin Heidelberg, 2002, (Lecture Notes in Computer Science, v. 2298). p. 241–253. ISBN 978-3-540-43678-2. Available at: <http://dx.doi.org/10.1007/3-540-47873-6_26>.

WANDERLEY, M. et al. The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, Taylor & Francis, v. 34, n. 1, p. 97–113, 2005.

WANDERLEY, M. M.; DEPALLE, P. Gesturally-controlled digital audio effects. In: CITESEER. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*. [S.l.], 2001.

WIDMER, G.; GOEBL, W. Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, Taylor & Francis, v. 33, n. 3, p. 203–216, 2004.

WING, A. M. et al. Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, The Royal Society, v. 11, n. 93, p. 20131125, 2014.

WOLFGANG, P. Perception and action planning. *European Journal of Cognitive Psychology*, v. 9, n. 2, p. 129–154, 1997. Available at: <http://dx.doi.org/10.1080/713752551>.

WÖLLNER, C. Self-recognition of highly skilled actions: a study of orchestral conductors. *Consciousness and cognition*, Elsevier, v. 21, n. 3, p. 1311–1321, 2012.

WöLLNER, C.; CAñAL-BRULAND, R. Keeping an eye on the violinist: motor experts show superior timing consistency in a visual perception task. *Psychological Research*, Springer-Verlag, v. 74, n. 6, p. 579–585, 2010. ISSN 0340-0727. Available at: <http://dx.doi.org/10.1007/s00426-010-0280-9>.

WöLLNER, C.; WILLIAMON, A. An exploratory study of the role of performance feedback and musical imagery in piano playing. *Research Studies in Music Education*, v. 29, n. 1, p. 39–54, 2007. Available at: <http://rsm.sagepub.com/content/29/1/39.abstract>.

ZERA, J.; GREEN, D. M. Detecting temporal onset and offset asynchrony in multicomponent complexes. *The Journal of the Acoustical Society of America*, v. 93, n. 2, p. 1038–1052, 1993. Available at: <http://dx.doi.org-/10.1121/1.405552>.