Computational Modeling of Synchronization and Leadership Dynamics in Small Group Interactions

by Sanket Rajeev Sabharwal

Department of Computer Science, Bioengineering, Robotics and System Engineering (DIBRIS)
University of Genova

Supervisors
Antonio Camurri, Gualtiero Volpe & Peter E. Keller

Ph.D. Thesis in Computer Science and Systems Engineering Computer Science Curriculum

October 31, 2023

Ph.D. Thesis in Computer Science and Systems Engineering

S.S.D. ING-INF/05, INF/01

Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi

Università di Genova, Via Opera Pia 13, I-16145 Genova, Italy

Candidate

Sanket Rajeev Sabharwal, sabharwal@edu.unige.it

Title

Computational Modeling of Synchronization and Leadership Dynamics
in Small Group Interactions

Advisors

Antonio Camurri, DIBRIS, Università di Genova, Italy

antonio.camurri@unige.it

Gualtiero Volpe, DIBRIS, Università di Genova, Italy

gualtiero.volpe@unige.it

Peter E. Keller, MIB, Aarhus University, Denmark

p.keller@clin.au.dk

External Reviewers

Sergio Canazza, Università degli Studi di Padova

canazza@dei.unipd.it

Alessandro D'Ausilio, Università degli studi di Ferrara

alessandro.dausilio@unife.it

Submitted On

October 31, 2023

Acknowledgements

With a heart full of gratitude, I dedicate this thesis to my late father, whose wisdom continues to be my guiding light. His teachings are the cornerstone of my life. For you words meant things, and that is an important lesson I learned from you. This scholarly journey commenced at the behest of your wish, which I embraced as my command. Your enduring desire for the well-being and advancement of your loved ones, children, and family fuels my drive. Thank you for instilling in me a relentless pursuit for the best.

I am eternally grateful to my mother, whose resilience and boundless love have been my stronghold. Her unwavering strength acts as the wind beneath my wings, propelling me forward through the enthralling dance of this Maya. To my brother, a paragon of strong character, you hold a special place in my heart. Neha, my better half, your patience, love, and encouragement always help me find ground.

Arriving in Italy marked the turn of a fresh page, yet I had little inkling then of the remarkable new chapter awaiting.Meeting Mary was the genesis of this new journey. She has been a nurturing presence, embodying the roles of a mentor, parent, guide, and cherished friend, a boon one could only seek from the divine. Durgesh, Yuriy, Rawad, Jamal, and Luca, your camaraderie has been a resonant echo of brotherhood, infusing this journey with layers of joy, support, and understanding. We often muse, “What is the point of reaching the top if you don’t have your people with you?” You transitioned from being the newfound friendships to the irreplaceable circle I hold dear, a bond that shall remain unswayed by geographical bounds. From yoga to vedanta, stoicism to muay thai, treks to dinners, debates to shared laughter, and through every shared embarrassment, your steadfast presence has become an integral verse in the narrative of my life.

The adage that your doctoral program hinges on your supervisors rings true, resonating profoundly with my experience. Heartfelt appreciation is extended to Professors Antonio, Gualtiero, and Peter, whose mentorship, unwavering support, and academic acumen have been instrumental in shaping this thesis. Their willingness to go above and beyond in creating a conducive work environment is something I hold in high regard. Antonio, your benevolent nature coupled with your proactive support through various challenges has been a source of encouragement. Your leadership, characterised by a clear conscience and work ethic, has been a source of inspiration and learning. Gualtiero, your composure in addressing my numerous inquiries and doubts provided a calming atmosphere for exploration and learning. Peter, your uniqueness is something I have acknowledged time and again - you truly are one of a kind. The substance of this thesis owe a significant debt to your invaluable input and guidance.

I find it imperative to acknowledge the profound teachings of eminent masters, philosophers, and authors, whose illustrious works have left an indelible mark on my ethos - instilling a steadfast belief in the mastery over one’s own destiny. The timeless wisdom in the Upanishads, Bhagavad Gita, Vedas, along with the seminal works of

Adi Shankaracharya, Swami Vivekananda, Ashtavakra, Swami Sarvapriyananda, not to mention the stoic doctrines of Marcus Aurelius, Seneca, and Epictetus, have etched a lasting imprint on my intellectual and spiritual canvas. These revered texts and individuals have not only enriched my understanding but have also been the moral compass guiding me personally. My experience of life, and particularly scholarly inquiries, would not be the same without them.

To all those mentioned, and the many unmentioned, who have touched my life, I extend my deepest gratitude.

Hari Om Tat Sat.

Dedicated to my treasured family, whose steadfast love and encouragement have been instrumental in my growth and development. Their support is my compass, guiding me in every endeavor.

My life's purpose is centered around reciprocating their faith and sacrifices. I remain unwaveringly committed to this objective, today and always.

Abstract

This doctoral thesis introduces *huSync*, a computational framework developed to assess non-verbal communication dynamics in small groups quantitatively. It aspires to contribute to Interaction Design and Human-Computer Interaction (HCI), focused on bridging computer science techniques, embodied design and somatics. *huSync* employs pose estimation algorithms to interpret movement trajectories from video sequences, offering a non-intrusive way to study entrainment in small-group settings based on established conceptual frameworks.

Joint actions in musical ensembles serve as the primary case study to explore how non-verbal body cues particularly influence interpersonal coordination and the directionality of information flow. These musical joint actions are exemplary instances of 'self-managed groups', illustrating the complex relationships between musical structure, entrainment dynamics, and mutual influence among ensemble members. These musical interactions are central to the research, serving as a universal language to understand nuanced human behaviours. Methods and results derived from three distinct studies on dyadic group dynamics are also presented.

Recent studies are making improvements in computationally modelling human behavior, sharing interesting techniques and approaches, and continues to remain an open area for research. *huSync* provides a

computational methodology and approach to model these behavioural mechanisms in small-group setups, and in this thesis, we present musical ensembles as a use case. By studying these subtle interactions in group settings, we begin to get a clearer perspective on how groups connect and interact. The potential of *huSync* lies in its capability to make these intangible elements tangible, offering a fresh perspective on the subtleties of small-group interactions.

huSync is versatile, managing diverse data sources and identifying essential movement attributes characteristic of group interactions, combining multi-modal signals, feature extraction, entrainment measurement, and analysis validation. It extends its applications to healthcare projects, emphasising music's universal role in non-verbal communication and interdisciplinary studies by integrating technology with economics, psychology, and the arts. It aims to bridge diverse disciplines, suggesting new paths for research in human movement sciences, especially regarding the use of markerless technologies in behaviorally driven computational research.

The findings from *huSync* are reliable and provide an alternative means for analysing human body movements, aiding in deepening the understanding of small-group dynamics and the elements contributing to successful collaborations. *huSync* seeks to provide dependable insights into evolving human behaviours by centralising the human body in small-group interaction-related contexts.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Non-verbal communication constitutes a foundational component of human interaction, serving pivotal roles in signifying meaning, conveying emotions, and establishing relationships [1]. Its importance becomes even more accentuated in specialised settings like musical ensembles, where reliance on non-verbal cues like body movements and facial expressions is fundamental for coordinating performances and achieving harmonious sound [2].

In various social contexts, synchronisation and directionality of influence, critical components of non-verbal communication, are instrumental in determining leadership dynamics [3; 4]. Due to their intrinsic reliance on non-verbal communication for synchronisation, musical ensembles present a distinct opportunity to study these multifaceted phenomena [5]. Recent studies have utilised motion capture and video analysis to investigate these facets in small-group setups such as musical ensembles [6; 7]. Nonetheless, a gap exists in understanding the relationship between these factors and leadership dynamics in small-group music performances. Standard methodologies, such as Motion Capture systems, involve attaching markers to performers and, because of this, have limited widespread application due to their intrusive nature. Fewer non-intrusive studies exist, and where they do, they often employ diverse methods, such as eye gaze tracking [8; 9]. Recent advances in deep learning have led to the development of human pose estimation (HPE) algorithms, and using them for non-verbal communication remains a largely unexplored medium.

This thesis introduces *huSync*, a computational model and framework intended for the automated analysis of human body movements and associated movement qualities. It includes a structured set of components and protocols facilitating an objective and empirical exploration of non-verbal cues. It utilises HPE algorithms to extract postural and movement data, offering a non-intrusive alternative for studying a variety of small-group settings in naturalistic contexts. Recognising the importance of investigating interpersonal dynamics within groups in real-world environments, *huSync* has been thoughtfully designed to ensure that findings reflect routine group interactions. Utilising HPE methods, we ensure participants can exhibit their natural movements, a contrast to the constraints posed by MoCap's form-fitting attire.

With this aim, *huSync* is applied to standard video footage of professional musical groups, facilitating the exploration of their inherent movements and interactions. The structured nature of musical performances, where interactions are often pre-determined in musical scores across many cultures, provides an optimal environment to probe non-verbal communication. Building on this premise, *huSync* is used to analyse and assess how participants' movements evolve throughout structures specific to musical scores.

We address research questions on the effects of musical structural features on ensemble coordination, directionality of influence, and leadership dynamics, thereby elucidating their complex interrelationship. Evidence suggests that synchronisation and directionality of influence critically shape the collective performance and potentially hint at leadership dynamics [10; 11]. We operationalise the *huSync* model and implement our computational methods to quanify and measure dyadic synchronisation and the directionality of influence within musical ensembles, thus empirically testing our hypotheses.

This thesis addresses important research questions about ensemble coordination, directionality, and leadership by combining the fields of embodied interaction design, computer science, and Human-Computer Interaction (HCI). It provides fresh insights into the dynamics of nonverbal communication within musical ensembles. The results have considerable implications beyond of music, even though the *huSync* framework is specifically designed for this musical setting. Musical ensembles, being self-regulating teams striving for impeccable performances under systematic structures defined by objective conventions like musical scores, provide an unparalleled lens for understanding nonverbal communication in broader group activities [12]. As we employ video-based pose estimation to analyse extensive real-world performance datasets across cultures and ensemble sizes, we magnify music's potential as a medium to dissect and elucidate the intricate communication dynamics inherent to diverse social groups [10; 11].

1.1 Objectives

Two primary objectives guide this thesis. First, the development of *huSync* (Human Sync), a computational framework tailored for automated entrainment measures analysis, particularly interpersonal synchronisation and directionality of influence in small groups. While traditional methodologies have certain advantages, they pose challenges for observing naturalistic behaviours. By leveraging pose-estimation algorithms, *huSync* non-intrusively extracts relevant data from video, paving the way for analysing synchronisation and directionality under varied leadership scenarios in realistic settings.

The need for *huSync* emerges from the requirement to study interpersonal coordination in environments reflective of everyday joint actions. Within the musical ensemble environment, dictated by scores, a suitable setting to analyse non-verbal communication is presented, focusing on dyadic pairs.

The secondary objective is to operationalise *huSync* for quantifying interpersonal coupling (dyadic) and directionality of influence amongst co-performers in a musical ensemble. Through the extraction of postural information from standard concert video recordings, we examine the relationship between musical texture and leader-follower relations. This not only helps us better understand the correlation between performers' movements and musical structures but also assesses the influence of these structures on ensemble coordination and leadership dynamics. By testing *huSync*'s robustness, we empirically contribute to the literature of dyadic mechanisms in small groups.

1.2 Research Questions and Hypotheses

This thesis addresses research questions related to nonverbal communication, interpersonal synchronisation, and leadership dynamics within musical ensembles:

1. **RQ1** How might computational methods reliably quantify interpersonal synchronisation in small-group interactions?

2. **RQ2** How do structural elements of a musical composition, including texture and phrase position, modulate interpersonal synchronisation patterns and relationships among co-performers?

3. **RQ3** How does the presence of leadership, whether clear or ambiguous, alter synchronisation and ensemble interaction dynamics?

4. **RQ4** Are there identifiable patterns of directional influence between melody instruments and their accompaniments within an ensemble?

5. **RQ5** Can statistical and computational techniques such as phase-locking values, Granger Causality, and human pose estimation algorithms offer a comprehensive, objective assessment of nonverbal cues, synchronisation, and leadership dynamics in small group contexts?

In pursuit of answers to these research questions, the subsequent hypotheses will be explored:

1. **H1** Computational and statistical techniques, such as phase-locking values and Granger Causality, will yield credible indicators of interpersonal synchronisation and leadership dynamics in small-group interactions.

2. **H2** Ensembles performing musical textures with ambiguous leadership (polyphonic) will exhibit heightened synchronisation compared to those with a clearly defined leader (homophonic).

3. **H3** Points of structural transition in the music, such as phrase endings, that raise coordination demands will increase synchronisation.

4. **H4** Granger Causality analysis is expected to show a more pronounced directional influence from melody instruments to accompaniment than vice versa, especially in homophonic textures.

Through this investigation, the thesis aims to empirically validate the proposed computational model and showcase its potential to address research questions on the dynamics of nonverbal communication, entrainment, and evolving leadership in musical groups. This knowledge base can potentially be extrapolated to diverse small team dynamics.

1.3 Thesis Outline

This The structure of this thesis is as follows:

1. Introduction (Chapter 1): Outlines the foundational objectives, research questions, and hypotheses guiding this study. Additionally, it points to relevant publications and introduces the datasets used.

2. Literature Review (Chapter 2): We review pertinent literature on non-verbal communication in musical settings, synchronization concepts, and human pose estimation. It provides a context within which this research is situated.

3. Methodology (Chapter 3): The *huSync* framework is detailed, and its procedural elements are articulated.

4. Dataset and Experiment Design (Chapter 4): We share details on the dataset being utilised, and the procedures involved in their annotation and segmentation steps. Additionally, it elucidates the experimental design for our studies.

5. Results (Chapter 5): Findings from the conducted studies are presented in this chapter. Each subsection provides a structured overview of the results from the respective studies.

6. Beyond Core Objectives (Chapter 6): During the course of this thesis we investigate additional areas where the methodologies and insights from this research found application, albeit outside the primary focus of the thesis.

7. Discussion (Chapter 7): We share our interpretations of the results and discuss their validity compared to existing literature, drawing tentative similarities and links and sharing the various limitations in our approach.

8. Conclusions (Chapter 8): We share our conclusive remarks, and suggest possible directions for future inquiry in this domain.

9. Appendix (Appendix A): Tables concerning the dataset for which experiments have been conducted, and the results obtained, have been organised in their respective sections.

1.4 List of publications

Results from the work carried out in this thesis were published in high-ranking peer-reviewed journals and international conferences, and their particulars have been systematically explained. Below, they are categorised according to their alignment with the research objectives. The first three studies address synchronisation, directionality of influence, and leadership dynamics within small-group setups, in particular musical ensembles. Insights from these studies laid the groundwork for the fourth, highlighting our research's interconnectedness and suggesting a promising avenue for further exploration. For clarity, individual studies within this document will bear titles such as 'Study 1', 'Study 2', and so on, serving as reference points for readers.

1. **Core Objectives**

   - **Study 1**: S. R. Sabharwal, M. Varlet, M. Breaden, G. Volpe, A. Camurri and P. E. Keller. *huSync - A Model and System for the Measure of Synchronisation in Small Groups: A Case Study on Musical Joint Action*, in IEEE Access, vol. 10, pp. 92357-92372, 2022 [13].

   - **Study 2**: S. R. Sabharwal, Arianna Musso, Matthew Breaden, Eva Riccomagno, Antonio Camurri, Peter E. Keller. *Analysing directionality of influence among ensemble musicians using Granger Causality*, in

International Conference of Kansei Engineering and Emotion Research (KEER), Barcelona, Spain, 2022 [14].

- **Study 3**: S. R. Sabharwal, M. Breaden, G. Volpe, A. Camurri and P. E. Keller. *Leadership Dynamics in Musical Groups: Quantifying Effects of Musical Structure on Directionality of Influence in Concert Performance Videos*. Currently under review [15].

2. **Beyond central objectives**

   - **Study 4**: A. Bergsland∗ and S. Rajeev Sabharwal∗. *Examining the Correlation Between Dance and Electroacoustic Music Phrases: A Pilot Study*, in Proceedings of the 18th International Audio Mostly Conference (AM '23). Association for Computing Machinery, New York, NY, USA. [16]

∗ These authors contributed equally.

1.5 Software repository and datasets

Following is a list of the software contributions, and the datasets utilised, for the studies conducted during the course of this thesis:

1. **huSync - Dyadic Synchronization**: This repository contains the code for the huSync system and framework, and also the data utilised for Study 1.

2. **Directionality of Influence**: This repository contains the code and data that was Utilised for the experiments performed in Studies 2 and 3.

3. **Correlation Music and Movement - Code Repository**: Code repository for the experiments conducted for the submission to AudioMostly 2023.

4. **Correlation Music and Movement - Online Material**: Online material that was a part of the submission for AudioMostly 2023. Contains the tables, references, and mocap example.

5. **MECS-Py**: A pythonic implementation of the Multi-Event-Class Synchronization (MECS) algorithm that is being utilised for investigating saliency of movements as part of future works.

Within these repositories, where applicable, the relevant datasets have also been uploaded for easy reproducibility for researchers.

Chapter 2

Literature Review

In HCI, the increasing integration of sensors and sensing technologies has driven a profound shift towards embodied interaction design. This paradigm acknowledges the body's central role in interface design and interaction, emphasising non-verbal communication as pivotal. The movement towards embodied interactions is not merely a consequence of advancing technology but also a reflection of recognising the body's significance in conveying subtle yet crucial information—primarily through somatic states. As software development for interface design evolves, it naturally demands a keen understanding of these non-verbal cues.

Musical ensembles, with their nuanced non-verbal exchanges and patterns of synchrony and influence, serve as an ideal setting to investigate into embodied interactions. However, the dynamics inherent to these ensembles, while fundamental, are multifaceted and intricate to decipher.

In this literature review, we discuss studies, perspectives, theories and bodies of work that are relevant to our research questions (see 1.2). In this process, we identify shortcomings in previous research, and make an attempt in this work to address the gaps. The aim is to build upon prior studies, setting the stage for the subsequent chapters of this thesis. We cover the following:

1. Non-verbal Communication
2. Synchronisation in Small Groups
3. Directionality of Influence and Leadership Dynamics
4. Computational Models for Studying Non-verbal Communication
5. Human Pose Estimation

2.1 Non-verbal Communication

Non-verbal communication includes a spectrum of cues, including facial expressions, vocal cues, touch, proxemics (the study of personal space), gaze, physical attractiveness, facial morphology, behavioural choices such as hairstyle, clothing, adornment, and appearance, and even material objects serving communicative functions within a society [17].

The scientific study of nonverbal behaviour traces its origins to Darwin's seminal work, 'On the Expressions of the Emotions in Man and Animal' [18], where he highlighted the evolution and adaptiveness of emotional expressions in animals and humans. Anthropological contributions, especially in kinesics (the study of body movement) [19] and proxemics [20], played a pivotal role in its development.

Nonverbal communication has garnered extensive interest across disciplines such as psychology, linguistics, medicine, sociology, anthropology, ethology, and law. Research in this domain examines the communication of emotional states in humans and animals, focusing on emotions and expressive features characterising specific states, such as 'the loud voice of extraversion' [21] and its implications for initial impressions.

Facial expressions, especially those conveying emotions, receive substantial attention in literature. Paralinguistic aspects such as voice quality, gestures, and the recently renewed focus on gaze are also under scrutiny. Nevertheless, dimensions of nonverbal communication beyond facial expressions remain relatively underexplored [17].

Nonverbal communication plays a pivotal role in social interactions. Within dyadic settings, a significant phenomenon is behavioural synchronisation. As interactions progress, participants converge in speech characteristics like loudness and speed, linked to rapport [22]. This behavioural synchronisation, also known as the 'chameleon effect,' underscores shared nonverbal behaviours such as foot tapping and face touching [23], fostering affiliation and often referred to as 'social glue' [24].

Another phenomenon, facial or emotional mimicry, involves imitating emotional behaviour [25]. Often perceived as a form of affective empathy or a 'low road' in empathy, mimicry aids in comprehending others' emotions [26], and its suppression can lead to reduced decoding accuracy in specific contexts [25].

Nonverbal communication is a complex aspect of our social interactions and profoundly shapes our behaviours. While facial expressions receive emphasis, voice quality, gestures, and gaze are equally pivotal in conveying information and facilitating social interaction. The roles of behavioural synchronisation and facial mimicry underscore nonverbal communication's significance in fostering affiliation and understanding emotions. A comprehensive grasp of nonverbal communication is essential for comprehending human social interaction [17].

2.1.1 Non-verbal Communication in Musical Ensembles

Effective nonverbal communication is essential for a successful musical ensemble performance. Instrumental movements and other physical gestures are crucial in sound production and conveying expressive intentions. When combined with auditory signals, these visual cues reveal the hierarchical structure of the music and enhance the overall performance quality. [27; 28].

Musicians in ensembles must achieve a balance between precise interpersonal coordination and the requisite flexibility for expressive renditions [29; 30]. This balance often manifests as spontaneous or premeditated expressive variations. These variations, seen in attributes like micro-timing deviations and local tempo changes, convey musical structure and stylistic interpretation [31; 32; 33; 34]. Ensemble performance necessitates individual variations and coordinated changes among participants [35].

The spectrum of non-verbal behaviours in musical settings spans complex body movements, nuanced facial expressions, intentional eye contact, and explicit gestures [28; 36]. Body movements unintentionally echo rhythm, anticipating the actions of fellow performers [28], while facial expressions enrich music's interpretative depth by conveying a diverse range of emotions [36]. Effective use of eye contact, a non-verbal tool, serves multifaceted roles within ensembles, including signalling transitions, aiding synchronisation, and reinforcing mutual understanding [28]. Gestures, mainly when employed by conductors, play a crucial role in establishing rhythmic frameworks [36]. In musical ensembles, it extends beyond visible gestures, including auditory signals and movements essential for sound production [28]. The actions of pianists, violinists, wind instrument players, and others are intrinsic to ensemble communication. Ancillary movements, like head nods or body sways, determine performance tempo and offer fundamental visual cues for interpersonal coordination.

Expressive variations in attributes like timing add a layer of non-verbal communication [28]. Ensemble musicians rely on shared mental representations of music developed through rehearsals, fostering alignment of performance goals [35]. Performance cues, guiding musicians and acting as landmarks in compositions [37; 38], facilitate coordination through established memory processes [39], ensuring alignment with shared objectives for synchronised performance. Effective non-verbal communication contributes to the coherence observed in musical ensembles, shaping collective musical experiences. Studies have shown how a comprehensive understanding of these dynamics provides insights into collaborative mechanisms in ensemble performances [40; 41].

We explore the following roles of non-verbal communication in influencing performance outcomes:

1. **Coordination and Synchronisation**: Non-verbal communication catalyses temporal and expressive alignment among ensemble members, ensuring coherent and well-coordinated performances [40; 42]. Observing fellow musicians' movements, expressions, and gestures enables real-time adaptation, fostering synchronisation and unity [43].

2. **Emotional Expression**: Facial expressions, body movements, and other cues convey music's emotional content, enhancing performance quality [44; 45]. Musicians communicate their unique interpretations, engaging audiences and peers in an emotional journey [46].

3. **Leadership and Directionality of Influence**: Non-verbal communication establishes and maintains leadership dynamics within ensembles [29; 35], as leaders effectively employ gestures and eye contact to guide and influence performances [6; 47]. Musicians align their performances with the leader's intentions, achieving a cohesive execution [48; 49].

Nonverbal communication in musical ensembles involves a complex interplay of visual and auditory cues, shared representations, and performance goals [27; 28; 36]. It necessitates coordination flexibility and is pivotal for cohesive and expressive performances [29; 31; 33]. Extensive research supports these insights, guiding the methodology and experimental design in this PhD Thesis [32; 34; 37; 38; 39].

2.2 Synchronisation in Small Groups

Synchronisation, central to group dynamics, is thoroughly explored across disciplines, including psychology, sociology, and musicology [50]. Beyond mere temporal alignment, it solidifies collective goals and social cohesion in small groups, such as musical ensembles [35]. This process manifests in behavioural, emotional, and cognitive forms.

Behavioural synchronisation pertains to coordinated physical actions like congruent movements [42; 51]. Emotional synchronisation encompasses shared emotional states within groups [52], while cognitive synchronisation aligns mental processes, including shared attention [53; 54].

Various theories elucidate synchronisation's mechanisms. Entrainment theory suggests individuals naturally align their actions with their environment's rhythmic patterns [55]. Conversely, intentional coordination posits conscious synchronisation to optimise group performance [2].

Synchronisation's degree in groups is influenced by factors such as individual differences, group structure, and task-specific characteristics. For instance, individual traits and experiences influence synchronisation propensity [29; 35], while group size and diversity affect synchronised behaviour emergence [42; 56]. Task intricacies, like complexity, also play a role [57; 58].

To visualise behavioural synchronisation, think of a rowing team aligning their strokes for efficient movement, as highlighted in Fig. 2.1a. Similarly, an orchestra requires behavioural and cognitive synchronisation, ensuring every musician plays in harmony, as depicted in Fig. 2.1b. In the case of cognitive synchronisation, J Buck pointed out a fascinating example from nature, where he hypothesised that fireflies possess a neural mechanism that automatically makes them synchronise the rhythmic flashing of conspecific males [59].

In small groups, synchronisation complexities, intertwined with leadership dynamics, enrich our understanding of effective collaboration [60]. For example, a lead musician may adjust their tempo to harmonise with a colleague—an instance of behavioural synchronisation influenced by leadership dynamics [6]. Similarly, one can observe military officials coordinate their movements in perfect harmony during parades 2.1g. Such moments emphasise the nuanced interplay of social interaction within groups.

Musical ensembles underscore synchronisation's importance, where performer coordination is imperative for success [42]. Here, synchronisation coordinates notes, phrases, and structural elements, influencing temporal accuracy and overall quality [6; 35].

We can parallel synchronisation's interplay with leadership dynamics in other contexts. A project manager directing a team is akin to a conductor guiding an orchestra, with synchronisation reflecting their leadership efficacy. These understandings have applications in enhancing group performance in classrooms, workplaces, or sports teams. Additionally, insights can guide technology designs like collaborative software and human-interactive robotic systems [50].

Understanding synchronisation allows for tailored strategies to enhance group performance, from training programs to foster synchronisation skills to policies encouraging diversity within teams [56]. This research area extends beyond academia, offering practical educational, business, and sports implications.

In technology, synchronisation principles can augment human-computer interaction. For instance, understanding user activity synchronisation could influence collaborative software tool designs or enhance interactions with social robots, as illustrated in Fig. 2.1e. Examining larger structures, insights into synchronisation can elucidate dynamics within communities, societies, and nations, emphasising the significance of synchronisation in fostering collective action.

The study of synchronisation in small groups offers profound insights into human social interaction. We gain invaluable knowledge applicable in diverse social and professional settings by investigating individual alignment within group contexts and influences such as leadership dynamics. This evolving research domain holds the potential to unveil methods to improve group performance, leadership efficacy, and societal cohesion.

2.2.1 What is Interpersonal Synchrony?

When discussing "Synchronisation in Small Groups", we often refer to how a collective group behaves in unity or harmony, perhaps in achieving a common goal or functioning as a unit, and it can include coordinated actions, shared emotions, or aligned objectives among all members.

In contrast, "Interpersonal Synchrony" narrows the scope from the group to individual-to-individual interactions, and it refers to the dynamic alignment and temporal coordination. Interpersonal synchrony refers to the dynamic alignment and temporal coordination between individuals [62; 63]. This complex phenomenon goes beyond mere imitation or mirroring to involve the precise timing and reciprocal adaptation of actions, gestures, and nonverbal signals between partners [64]. This kind of synchrony is not just about doing the same thing simultaneously; it is about how two people can anticipate, mirror, and reciprocate each other's actions, emotions, and intentions in real time.

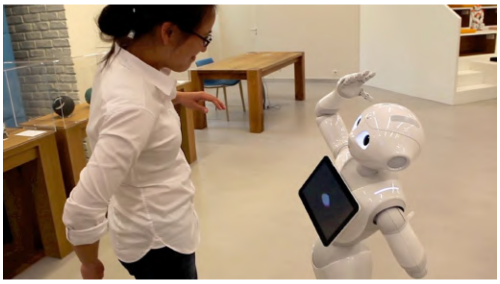
The study of interpersonal synchrony originated in developmental psychology research on caregiver-infant interactions, where attunement and synchrony in gaze, affect, and vocal rhythms were deemed crucial for secure attachment and social development [65; 66]. However, researchers found that interpersonal synchrony subsequently facilitated social connection and task performance in contexts beyond infancy [67; 68]. For instance, synchrony has increased affiliation, rapport, and prosocial behaviour between adults [58; 69].

(a) Rowing team demonstrating behavioural synchronisation.

(b) Orchestra illustrating synchronised performance.

(c) Ensemble showing synchronisation and leadership.

(d) Dance troupe in harmonised performance.



(e) Pepper robot interacting with child [61].

(f) Fireflies exhibiting synchronised flashing.

(g) Military officers in aligned formation.

In recent years, interpersonal synchrony has become an active multidisciplinary research area spanning social signal processing, human-robot interaction, music psychology, and clinical psychiatry [64; 70]. This broad interest reflects a growing recognition of the role of precise temporal coordination in social interaction. However, methodological barriers have constrained progress, especially the need for automated tools for measuring synchrony dynamics in natural settings [71]. Traditionally, developmental psychologists examined interpersonal synchrony through manual coding of videotaped interactions to mark gaze patterns, affect sharing, and contingent responsiveness [65; 66]. However, such labour-intensive analysis cannot scale to larger groups or capture subtle temporal nuances. Computational methods are needed to achieve an objective, quantitative perspective.

Critical elements for the emergence of interpersonal synchrony identified by psychology research include sustained mutual focus, prompt responsiveness, matching of activity and energy levels, and attunement to the partner's state [65; 72]. Underpinning this behavioural alignment are shared mental models, allowing partners to anticipate actions and coordinate adaptively [73]. Developing computational metrics that capture these dynamic elements poses an interdisciplinary challenge involving signal processing, machine learning, and construct grounding in behavioural theory.

Researchers made initial progress, using motion energy analysis (MEA) to quantify body movement synchrony in psychotherapy [74] and recurrence analysis to measure postural coordination in conversations [62]. However, substantial scope remains for sophisticated automated synchrony measures to unlock new insights into social interaction patterns and deficits in conditions like autism. Advancing methods to assess interpersonal synchrony can enable diverse applied domains while addressing basic theoretical questions about temporal coordination in human relations.

2.2.2 Interpersonal Synchrony in Musical Ensembles

Interpersonal synchrony in musical ensembles is a fascinating area of study that intersects the fields of music, psychology, and neuroscience. This phenomenon refers to the temporal coordination between individuals during a musical performance, creating a unified and harmonious output. The study of interpersonal synchrony in musical ensembles provides insights into the complex interplay of individual skills, group dynamics, and the overarching musical structure [64].

Interpersonal synchrony in musical ensembles manifests the broader concept of synchrony in human interactions. Synchrony generally refers to the temporal coordination during social interactions, requiring the perception and integration of multimodal communicative signals and continuous adaptation [64]. This concept has been studied extensively in various contexts, including early development, language learning, and social connection.

Interpersonal synchrony in musical ensembles is crucial for cohesion in performances. It involves coordinating various elements, including rhythm, tempo, dynamics, and articulation. The musicians in an ensemble must continuously adapt their performance in response to the cues from their fellow musicians, thereby achieving a high level of synchrony [75].

Several studies have investigated mechanisms underlying interpersonal synchrony in musical ensembles. For instance, Varni et al. [76] conducted a study on "Emotional Entrainment in Music Performance", highlighting emotional synchrony's role in achieving a unified performance. Similarly, another study by the same authors focused on "Measuring Entrainment in Small Groups of Musicians", providing insights into the dynamics of interpersonal synchrony in small musical ensembles [77].

The study of interpersonal synchrony in musical ensembles is broader than just the coordination of musical elements. It also involves the coordination of non-verbal cues, such as body movements and facial expressions. For instance, a study by Richardson, Marsh, and Schmidt [78] explored the effects of visual and verbal interaction on unintentional interpersonal coordination, highlighting the role of non-verbal cues in achieving synchrony.

Interpersonal synchrony in musical ensembles is a complex and multifaceted phenomenon that involves the coordination of musical elements and non-verbal cues. It requires a high level of individual skill and group dynamics, making it a fascinating study area for music, psychology, and neuroscience researchers.

2.3 Directionality of Influence and Leadership Dynamics

Directionality of influence is a crucial concept when examining leadership dynamics in small groups, as it pertains to the process through which specific individuals exert a more substantial influence on others, consequently shaping group decisions and behaviour [79]. In this section, we will explore the existing literature on the directionality of influence and leadership dynamics, highlighting the primary findings and insights from prior research.

Leadership dynamics involve the patterns and processes by which leaders emerge, exercise influence, and interact with other group members. An extensive body of research has investigated the role of formal and informal leaders in small groups, revealing that leadership can emerge from various sources, such as individual expertise, personal charisma, and social connections [80]. The notion of shared or distributed leadership has gained traction in recent years, suggesting that leadership roles and responsibilities can be fluid and distributed among group members, depending on the context and demands of the task [81; 82].

When studying the evolution of leadership dynamics in small groups, reviewing the direction of influence is important. Research has revealed that leader-follower interactions can be reciprocal, meaning that both parties can influence each other through ongoing feedback loops [83]. To measure this dynamic process, techniques like Granger causality (GC) analysis are used in neuroscience and psychology to determine the causal relationships between individuals and how they influence each other's actions or behaviour [84].

In musical ensembles, the directionality of influence and leadership dynamics play a pivotal role in shaping the quality and expressiveness of performances. The ensemble is a microcosm of a small group where leadership roles can shift fluidly based on factors such as the individual's instrument, position within the group, and expertise [6]. For instance, a conductor often serves as the central figure guiding the ensemble, strongly influencing the group's performance. However, the directionality of influence is not solely from the conductor to the musicians. Musicians, particularly section leaders or soloists, can also influence the group's performance, contributing to the overall sound and cohesion of the ensemble [85]. This dynamic interplay of influence underscores the complexity of leadership dynamics within small groups, illustrating how leadership can be a shared and reciprocal process. Studying musical ensembles offers a rich context to understand the nuances of directionality of influence and leadership dynamics in small groups. It provides insights we can extrapolate to other social and professional settings.

Researchers identified several factors that shape the directionality of influence and leadership dynamics in small groups. Personality traits, cognitive abilities, and emotional intelligence can affect a person's propensity to assume a leadership role or be influenced by others [86; 87]. Group factors, including group size, diversity, and familiarity among members, can also influence the emergence and stability of leadership roles and the directionality of influence [88].

The literature on the directionality of influence and leadership dynamics has yielded valuable insights into the factors contributing to the emergence of leadership roles, the dynamics of leader-follower interactions, and the impact of individual and group factors on these processes. This foundational knowledge serves as a basis for investigating the directionality of influence and its relationship with synchronisation in musical ensembles, thereby illuminating the underlying mechanisms that drive the complex interplay between leadership and group coordination.

2.4 Human Pose Estimation

Human pose estimation (HPE) is the computational task of detecting human body posture by estimating the location of critical joints or body parts from images or video. It has become a central topic in computer vision due to its diverse applications ranging from human-computer interaction to robotics and activity recognition. Following a brief literature review of the evolution of HPE, we discuss the potential of studying non-verbal communication.

2.4.1 Early Methods

Historically, human pose estimation (HPE) methods relied on manually created features and probabilistic graphical models, specifically pictorial structures [89]. These models depicted the human body as a collection of interconnected parts and utilised features such as HOG to identify body parts and determine the overall pose [90]. However, experts widely acknowledge that we must improve these methods to handle complex movements and obstructions.

With the advent of deep learning, the landscape of HPE transformed. Toshev and Szegedy [91] pioneered this change, using deep neural networks to regress body joint coordinates directly. Their approach dramatically outperformed existing techniques on the FLIC dataset. Subsequent innovations such as the Convolutional Pose Machines by Wei et al. [92] and the Stacked Hourglass network by Newell et al. [93] further refined joint detection, even in challenging scenarios.

2.4.2 Multi-Person Pose Estimation

Researchers designed initial methods for single-person pose estimation. However, challenges arose when applying them to real-world images containing multiple individuals, variations in appearance, occlusions, and an unpredictable number of people. Different strategies emerged to address this: top-down methods that first identified individual bounding boxes before estimating poses and bottom-up methods that identified joints and later grouped them into individuals.

Pioneering top-down solutions include proposals by Insafutdinov et al. [94] and He et al. [95]. Contrastingly, bottom-up methods like OpenPose [96] detected joint candidates and then organised them using part affinity fields. Recent endeavours have attempted a hybrid of these strategies for optimised multi-person pose estimation [97].

2.4.3 3D Human Pose Estimation

Transitioning from 2D to 3D pose estimation introduces complexities due to depth information loss. Previous methods leveraged 3D scans or multi-view images. Modern techniques integrate 2D joint estimates across time or impose bone length constraints [98]. Recent research is now exploring implicit depth learning and self-supervised methods that utilise unlabeled video [99; 100].

2.4.4 Applications of HPE

HPE's rising precision has spurred diverse applications such as gesture recognition and control interfaces in HCI, VR/AR avatars in gaming, detecting unusual activities in surveillance systems, understanding customer behaviour in the retail industry, predicting pedestrian or car movements in autonomous vehicles, enhancing athlete performance and monitoring patients in health sciences, and aiding people who are hard of hearing through sign language recognition. The growing accuracy of pose estimation methods is paving the way for broader adoption across industries. Lightweight models are also enabling real-time uses on edge devices. As algorithms improve and applications diversify, the exploitation of HPE in HCI-related applications becomes a promising research frontier.

2.4.5 Potential of HPE in studying Non-Verbal Communication

Human Pose Estimation (HPE) has witnessed immense growth and application in diverse fields, from sports analysis to medical support and surveillance [101; 102; 103]. Such interest stems from the capability of HPE to detect human body postures and movements from images and videos and inherently speaks to its potential as a dependable tool for non-verbal communication analysis.

A closer inspection reveals that HPE offers more than what meets the eye. It does not just recognise body parts or postures; it delves into the details of movement and position, often the silent yet profound indicators of socio-emotional states [104]. For instance, a person's defensive posture can reveal discomfort or apprehension, even if their words convey confidence. Moreover, tracking these nuanced movements over time can provide invaluable insights into behavioural mimicry, indicating rapport and trust or hinting at dominant-subordinate dynamics in group interactions.

Studies like those by Angelini et al. [105] showed the capability of posture-based algorithms in recognising diverse human actions. Similarly, Fanello et al. highlighted HPE's role in real-time activity recognition using RGBD images with a one-shot learning approach [106]. These findings emphasise the importance of postures and movements in non-verbal cues, which can lead to a deeper understanding of group dynamics, societal roles, and even disorder diagnosis based on movement patterns.

Still, complexities such as occlusion, object interaction, or motion blur can affect the reliability of pose estimation algorithms. Despite these challenges, the advancements in HPE, as evidenced by Osokin's work on real-time performance on edge devices [107], and the utilisation of OpenPose for diverse applications [96; 108; 109], demonstrate its growing robustness and reliability. Studies like that of Goyal et al. [110] attempt to address these obstacles using event-driven cameras. These advancements make the potential of HPE in decoding non-verbal nuances in human interactions promising.

With the increasing relevance of HPE in human interactions, researchers aim to harness its potential in non-verbal communication analysis. Still in its nascent stages, this area encourages the intersection of computer vision, behavioural sciences, and communication studies [111]. There is currently a limited body of research in non-verbal communication analysis, and this underscores the motivation for study to utilise such non-intrusive techniques to study interpersonal synchrony, directionality of influence, and leadership dynamics.

2.5 Computational Models for studying Non-verbal Communication

2.5.1 Measuring Interpersonal Synchrony

The study of interpersonal synchrony, defined as the dynamic and reciprocal adaptation of the temporal structure of behaviours between interactive partners, has been a topic of interest for researchers across various disciplines for many years [64]. This interest stems from recognising the critical role that non-verbal communication plays in human interaction, particularly in small-group settings where leadership dynamics are often at play [112; 113]. In such settings, the ability of group members to synchronise their behaviours and adapt to changes in the behaviour of others can strongly influence the group's overall performance. Such a phenomenon is particularly evident in musical ensembles, where the success of the performance relies heavily on the ability of the musicians to coordinate their actions and maintain harmony [2].

Researchers proposed several approaches over the years to analyse interpersonal synchrony. Historically, these methods involved the manual coding of non-verbal cues by trained human coders. While providing valuable insights into the dynamics of human interaction, this approach was labour-intensive and time-consuming, limiting the scope and scale of the studies that could be conducted [114]. Besides, manual coding methods were subject to various forms of bias and error, such as coder bias and inter-coder reliability issues, which could potentially affect the validity and reliability of the findings [64; 115]. Therefore, there was a clear need for more efficient and objective methods for studying interpersonal synchrony.

Addressing this, researchers began to explore using automated coding techniques for collecting nonverbal behavioural data. These techniques, which often involve advanced computer algorithms and machine learning methods, can collect data more effectively, objectively, and efficiently than traditional manual coding methods [64; 115]. These techniques have been particularly beneficial for studying interpersonal synchrony in the context of nonverbal communication and leadership dynamics in small-group settings. Researchers have made use of such advanced techniques to measure the degree of synchronisation and directionality of influence among the musicians [116], and also adaption and anticipation models of sensorimotor synchronisation (SMS) [4]. Newtson and colleagues developed one of the initial methods where they overlaid transparency on a video screen's still frame, tallied body part changes to create a time series, and evaluated movement changes over time [117; 118; 119]. Despite its innovativeness, this method was laborious, provided a coarse view of synchrony, and struggled to capture intricate, dynamic human interactions [120].

In contrast, MoCap methods excel in capturing and quantifying interpersonal synchrony in structured group activities, discerning subtleties that other techniques might miss. These methods provide insights into spatial proximity, motion information, and interactive behaviours prevalent within group members [121]. A study by Novotny and Bente used MoCap data to create computer animations of dyadic interactions, where participants performed a repetitive motor task in unison [122]. Observers assessed these animations derived from both partners' synchronised full-body motion capture data to gauge synchrony and leader-followership dynamics in the dyad. The study revealed a significant correlation between most synchrony measures and observer perceptions, with phase synchrony and Pearson correlations strongly associated with these observations. MoCap's precision in capturing fine and gross motor movements, such as finger movements and limb motions, respectively, is particularly beneficial for studies related to motor control conditions, including Parkinson's Disease [123]. Overall, MoCap offers a detailed and comprehensive perspective on interpersonal synchrony across diverse settings.

While MoCap systems are known for their high-frequency data capture, accuracy, and low noise levels, they come with challenges. These specialised systems bear a significant cost and introduce methodological complexities. One particular challenge researchers often encounter is restricting natural movement due to the tight-fitting suits often required for motion-tracking [124]. Recognising these constraints, the research community has focused on alternative solutions, specifically video-based tracking methods. Researchers have recognised two major approaches: pixel or frame differencing methods and human pose estimation algorithms. Pixel-based or frame differencing methods capture motion in a specific region of interest [74; 125; 126]. Studies like Hadjakos et al. used the Kinect camera to analyse nuances in head movement for studying synchronisation of a violin duet performance [127]. On the other hand, human pose estimation algorithms track human body key points, and algorithms such as OpenPose, AlphaPose, or OpenFace have recently gained traction [96; 128; 129]. OpenFace, an open-source facial behaviour analysis toolkit, has been used in analysing nonverbal cues such as facial expressions, eye gaze, and head movements from video recordings. Similarly, OpenPose's real-time multi-person pose estimation has shown its efficacy in studying nonverbal communication, presenting results that stand promisingly alongside MEA [130].

Transitioning from MoCap to video-based methods has enabled studies in naturalistic contexts, offering efficient and less restrictive tools for analysing human motion and interaction. As research shifted from manual coding to advanced computer-based techniques, the study of interpersonal synchrony has become more streamlined. These techniques have paved the way for deeper exploration of human interaction and group dynamics previously challenging to investigate.

These methodological advancements provide objective insights into leadership dynamics within small groups, informing the formulation of strategies to optimise group performance.

2.5.2 Measuring Directionality of Influence

The dynamic interplay of sounds and movements in musical performances coordinates over varying timescales. Sounds align on a millisecond level, while body movements correspond to higher levels of musical structure, such as phrases[131; 132; 133]. This coordination is not static but evolves, reflecting the dynamic nature of musical communication [43; 116; 134].

In musical ensembles, performers often assume varying roles, with some leading and others following. These roles can be explicitly assigned or emerge spontaneously due to task structure or individual characteristics[29; 36; 135; 136]. Optical motion capture systems have been used to study these dynamics, revealing that leaders' movements often precede those of followers[42].

The study of the directionality of influence and information flow in group dynamics, particularly in musical ensembles, has seen the application of various computational approaches. One such approach is GC, a statistical concept used to determine whether one time series helps forecast another. GC has been used to quantify the flow of information between performers, revealing the influence of different performers and the dynamics of their interactions[7]. In a study by Klein et al. [137], GC was applied to measure information flow in sound during a musical performance. The study involved professional violinists playing along with recordings of two folk pieces. They compared the amplitude envelopes of their performances with those of the recordings using GC to measure information flow and cross-correlation to measure similarity and synchronisation. The study found that the measure of information flow was higher from the recordings to the performances than vice versa, and this decreased as the violinists became more familiar with the recordings over trials. This decline in information flow was consistent with a gradual shift from relying on auditory cues to predict the recording to relying on an internally-based (learned) model built through repetition. The study also found that the violinists became more synchronised with the recordings over trials. These results shed light on the planning and learning processes involved in aligning expressive intentions in group music performance and laid the groundwork for applying GC to investigate information flow through sound in more complex musical interactions. Researchers employed GC to predict a musician's upcoming movements based on the immediate past movements of fellow musicians [137].

Pertinent to our research, looking at four relevant previous studies on GC in ensembles, one analysed position [116] and three analysed acceleration [6; 47; 138]. Collectively, these investigations lay the groundwork for our exploration into ensemble dynamics and performer roles.

Bishop et al. [138] conducted a study to investigate the role of visual contact in the interaction between musicians during a duo performance. They analysed the head movements of pianists and clarinettists during performances of a contemporary piece, focusing on the effects of visual contact on movement coordination and interaction. The study found that visual contact between performers increased the consistency and coordination of their movements and that performers exchanged cueing gestures during held notes. The study also used GC to estimate the direction and magnitude of influence between performers' head movements. The results showed that leader-to-follower influence was more potent in sections of the piece where a melody/accompaniment structure implied leader/follower relationships. The study suggests that visual contact can enhance movement-based interaction between performers and may serve as a social motivator that encourages creative thinking and risk-taking during performance. The study used GC to estimate the influence of one performer's head movements on another by comparing two models. In the restricted model, they predicted observations of the second performer's movements only by lags of their movements. In the unrestricted model, they predicted them by lags of both their movements and the first performer's movements. They then compared the predictive ability of the two models to estimate influence likelihood.

In D'Ausilio et al. [6] study, GC was used to investigate the causal relationships among the movement kinematics of conductors and violinists during the execution of Mozart pieces (information flow between the baton and bow motion). The study aimed to explore whether conductors' kinematics were associated with a differential influence on musicians' performance (driving force) and if this was able to affect inter-musician interaction (interaction strength). The results showed that the increase of conductor-to-musician influence and the reduction of musician-to-musician coordination (an index of successful leadership) parallel the quality of execution, as assessed by musical experts' judgments. The research quantitatively displayed the GC pattern among conductors and musicians as a sensorimotor conversation between several individuals: musicians adjust their performance based on non-linguistic motor messages from other musicians and the conductors.

Chang et. al. [116] investigated the role of leadership and visual information in interpersonal coordination during musical ensemble performances. The study used GC analysis to examine the direction and magnitude of information flow among performers' body sway time series. The results showed that leadership assignment influenced information flow, with leaders having a more significant influence on followers' body sway than vice versa or between followers. Visual information also played a role, with the influence of leader-follower dynamics being higher in the visual present condition. The study also found that total interpersonal body sway coupling correlated with the performers' rated goodness of ensemble performance.

Hilt et al. [47] investigate the role of different channels of communication in ensemble music performance, specifically focusing on the interaction between the conductor and two sections of violinists. The study uses GC analysis to examine the causal relationships between the conductor's movements, the bow and head movements, and the synchronisation of the two sections of violinists. The study finds that the patterns of sensorimotor information carried by the bow and head movements are distinct, with the bow movements exhibiting a robust leader-follower relationship between the conductor and the violinists. The study also finds that the conductor's movements influence the synchronisation between the two sections of violinists and that the perturbation of the communication flow leads to changes in the causal relationships between the different movements and synchronisation. The study highlights the importance of multimodal communication in ensemble music performance and the potential of GC analysis to investigate complex non-verbal communication.

This chapter systematically reviews the literature pertinent to interpersonal synchrony, the directionality of influence, ensemble dynamics, and their underlying methodologies, shedding light on the progression from rudimentary techniques to advanced computational techniques. This comprehensive review intends to establish a foundational understanding of non-verbal communication in musical settings. The subsequent chapter will delineate the research methodology used to answer research questions as discussed in 1.2, built upon the insights extracted from this literature review.

Chapter 3

Methodology

This chapter presents the *huSync* computational model and how we utilise it to quantify and analyse interpersonal synchrony and the directionality of influence in small-group interactions. Drawing upon multi-modal signals such as audio and video, it offers a flexible and adaptable methodology that accommodates various processing techniques based on the experiment design to answer specific research questions.

Central to our approach is the concept of entrainment. We emphasise synchronisation measurement and the directionality of coupling among dyadic pairs. Specifically, *huSync* leverages phase locking values (PLV) to determine interpersonal dyadic synchronisation, while the Granger Causality (GC) method is employed to discern the directionality of coupling. This process aids in revealing the internal dynamics and influences within groups, providing a nuanced understanding of their collective interactions.

To underscore our methodology's practical and empirical value, we highlight its application and utility using performance recordings from the Omega Ensemble, a professional chamber music group. This helps showcase its empirical credibility using a real-world example. From its foundational theory to its tangible applications, the aim is to offer a replicable and robust model for studying small-group interactions and their dynamics.

3.1 Computational Model and Framework

To address the research questions in Section 1.2, we introduced *huSync*. This computational model quantifies entrainment in small group interactions, emphasising synchronisation and directionality of coupling within dyads. Grounded in a framework that evaluates expressiveness through body movements and gestures and building on prior research, *huSync* merges a multi-modal approach with rigorous statistical methods, offering a comprehensive tool for investigating small-group dynamics.

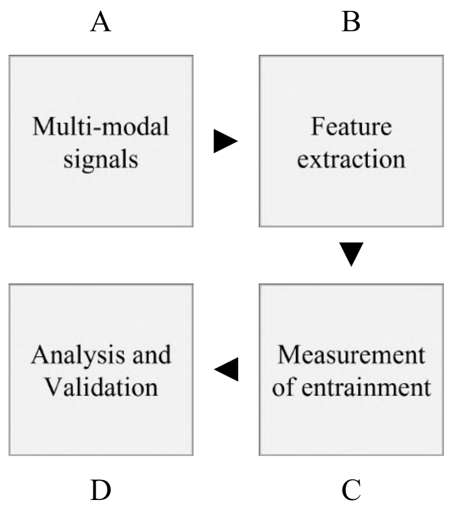


Figure 3.1: The huSync computational model and framework for entrainment measurement in small-group settings

Designed for adaptability, it can integrate alternate techniques tailored to specific applications at each stage. As illustrated in Fig. 3.1, the model comprises four primary blocks:

1. **Multi-modal signals (Block A)**: This captures diverse input signals like video, audio, physiological signals, and motion capture data for a holistic assessment of group coordination.

2. **Feature extraction (Block B)**: Input signals undergo pre-processing to glean features that elucidate bodily motion, acoustics, and interpersonal dynamics. For instance, video data are subjected to pose estimation algorithms to extract 2D or 3D skeletal joint coordinates. Audio data, on the other hand, yields features such as loudness, pitch, and timbre.

3. **Measurement of entrainment (Block C)**: We analyse the group's behaviour using extracted features in this section. We use metrics such as interpersonal synchronisation, calculated using the Phase Locking Values obtained from body sway patterns. Additionally, we determine directionality by performing Granger Causality analysis on time series data.

4. **Analysis and validation (Block D)**: Here, statistical techniques evaluate the impact of varying experimental conditions on group coordination. Cross-referencing measures from diverse modalities reinforce the validity of results.

3.2 Phase Locking Values and Granger Causality: Key Metrics in huSync

Entrainment measurement, particularly synchronisation and directionality of coupling among dyadic pairs, is at the core of our methodology. The *huSync* model utilises Phase Locking Values (PLVs) to quantify synchronisation and the GC Method to discern the directionality of coupling. By doing so, it maps influences within a group, illuminating the dynamics of collective performance.

We derive PLVs from the phase values of motion trajectory signals to ascertain the synchrony between two signals. We obtain these phase values for each frequency using the Fast Fourier Transform (FFT). Our study chooses relative phase values to evaluate dyadic synchronisation among co-performers, as outlined in a previous study [139]. We compute the PLVs by examining the phase difference between the head motion trajectories of paired co-performers.

$$PLV = \left| \frac{\sum_{t=1}^{n} e^{i(\Theta_1 - \Theta_2)}}{n} \right| \tag{3.1}$$

This equation's synchrony range is [0,1], with 1 indicating peak synchrony [140; 141].

GC determines whether one time series can predict another using linear regression modelling. Suppose variable X Granger causes variable Y; changes in X come before changes in Y. In that case, we seek evidence of this causality by regressing Y against its lagged values and the lagged values of X. If the coefficients on the lagged values of X prove significant, then X Granger-causes Y. We test the causality through regressions:

Autoregressive model of Y

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_n Y_{t-n} + e_t \tag{3.2}$$

Extended model with X

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_n Y_{t-n} + b_1 X_{t-1} + b_2 X_{t-2} + ... + b_n X_{t-n} + e_t \quad (3.3)$$

If coefficients b_1, ..., b_n in the extended equation significantly differ from zero, X Granger-causes Y, confirmed via F-test or chi-square test.

3.3 An Instance of huSync to Measure Dyadic Synchronisation Among Musicians

In Study 1, we introduced an instance of the *huSync* model and framework as depicted in Fig. 3.2. This illustration corresponds to the structure shown in Fig. 3.1. The process begins by selecting videos that meet our criteria. The first block (Fig. 3.2 (A)) captures both video and audio signals from standard recordings. The second block (Fig. 3.2 (B)) processes videos through multi-person pose estimation, detecting a participant's body key points frame by frame. It yields a JSON file that organises participants sequentially, each identified by key points. We use this trajectory data to extract kinematic information. In our specific use case, we are interested in the trajectory of the head and extracting the nose key point (key '0') to represent body sway. Simultaneously, we derive acoustic features, such as pulse clarity and event density, from the video's audio signals. In the third block (Fig. 3.2 (C)), we compute interpersonal synchronisation. Ultimately, the fourth block (Fig. 3.2 (D)) carries out statistical analysis and validates the PLV results in line with hypothesis-driven questions. We validate the interpretation robustness of primarily heterogeneous results cross-modally using acoustic features.

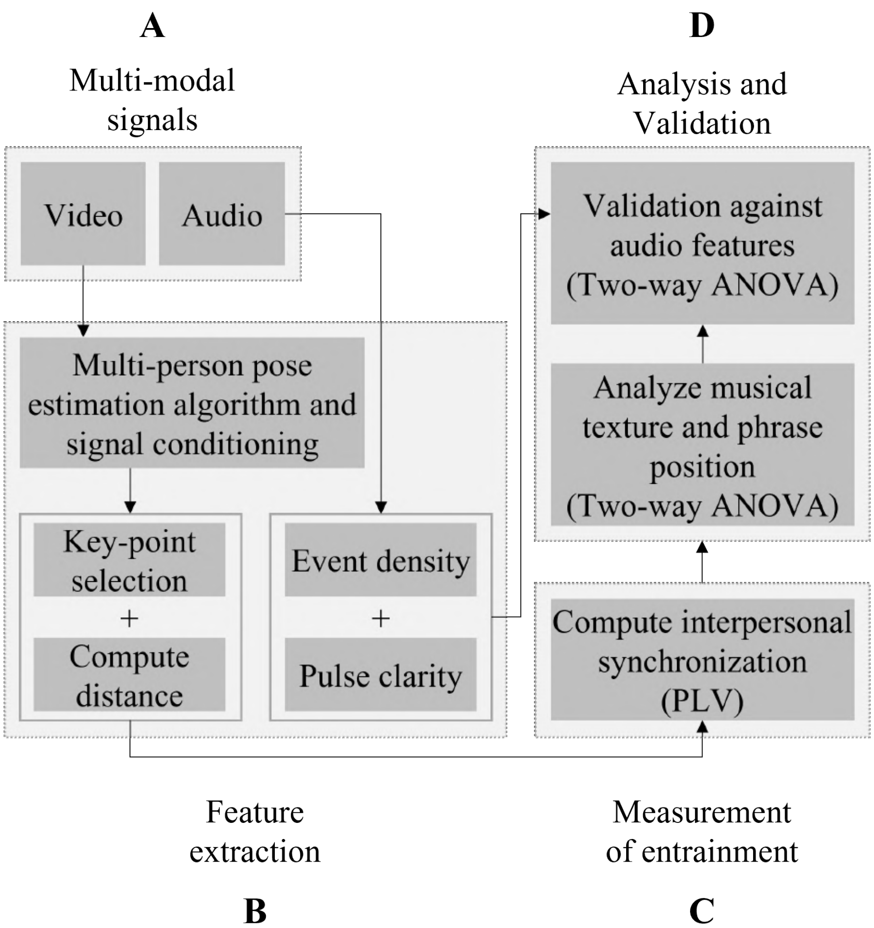


Figure 3.2: An instance of the computational framework and the huSync system architecture to compute dyadic synchronisation.

3.3.1 huSync Process Pipeline for Dyadic Synchronisation

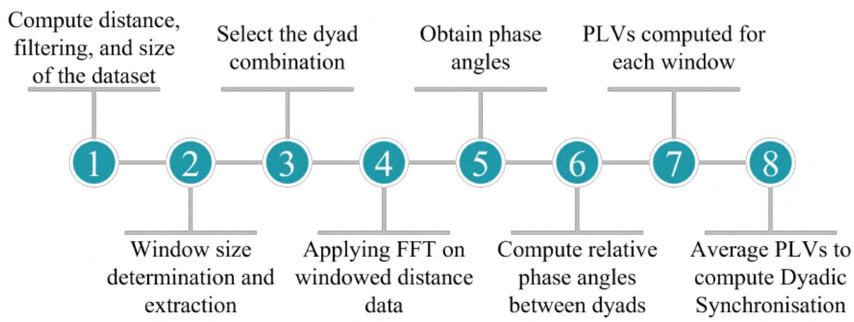


Figure 3.3: An illustration of the process pipeline for computing Dyadic Synchronisation.

huSync's follows an 8-step computational methodology, as showcased in Fig. 3.3. This pipeline integrates the elements of blocks B and C from Fig. 3.2, focusing on operations that manipulate data extracted from pre-processed JSON files to determine the final dyadic synchronisation in small-group contexts. Before leveraging the *huSync* model, we extract data from the JSON files, selecting the key point of interest based on the experimental arrangement.

Given its adaptability and flexibility, *huSync* can determine dyadic synchronisation across a complete video phrase or pinpoint specific segments, like the start, middle, and end, addressing the research inquiries outlined in section 1.2. Below, we provide details on these steps using a simulated dataset comprising 15 data points between two performers. Fig. 3.4 illustrates steps 1 to 6, while Fig. 3.5 showcases steps 7 and 8.

1. **Dataset Computation: Distance, Filtering, and Size:** For feature extraction, as depicted in Fig. 3.1 (C), the Euclidean distance between extracted nose key points, either single or multiple, is computed from raw (x,y) coordinates. Pose estimation on videos often introduces noise, so we assess the need for filtering. If required, the Savitzky-Golay filter, which

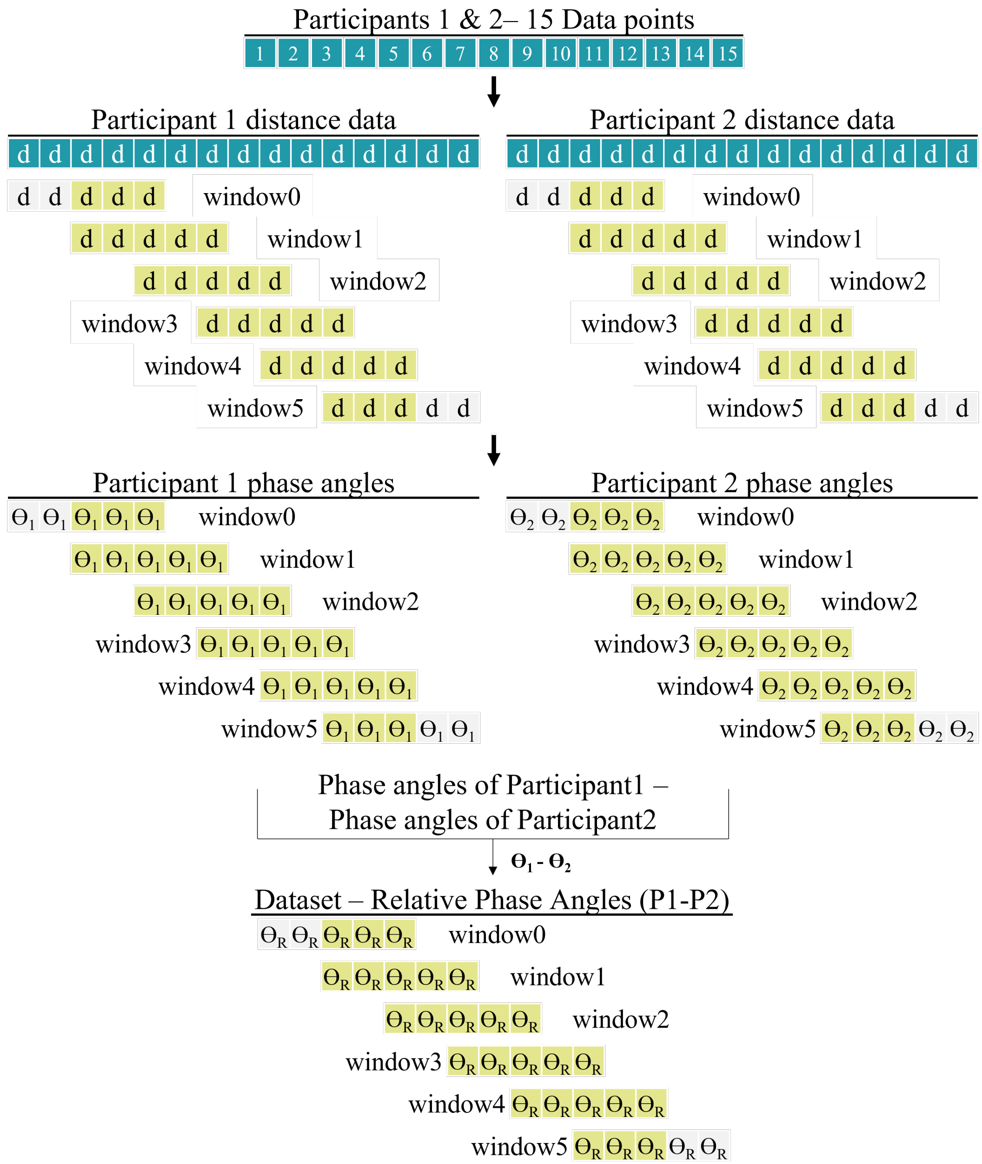


Figure 3.4: Simulated Example illustrating steps 1 to 6.

preserves signal features, is applied [142][143]. The dataset's size is determined to ensure synchronisation level analysis over desired time intervals.

2. **Window Size Determination and Extraction:** We adopt a sliding window approach to capture both local and global data trends. In our example, we set a window size of 5 and a step size of 2, resulting in 6 windows.

3. **Dyad Combination Selection:** Our focus is on situations involving two participants in our simulated example, commonly referred to as dyads. The *huSync* framework supports manual dyad selection or automated computation for all potential pairs within a larger group. For such dyad combinations, the calculation involves determining the number of ways n members can be paired.

   When forming pairs or dyads from a group of n individuals, the formula is akin to counting the number of edges in a fully connected graph with n nodes. Each node (or individual) in such a graph is connected to every other node. The total number of unique connections (or edges) can be simplified, especially when $r = 2$, to:

   $$C\left(n, 2\right) = \frac{n(n-1)}{2} \tag{3.4}$$

   This equation $\frac{n(n-1)}{2}$ provides a straightforward method to compute the number of dyads from n members. It calculates the cardinality of the edges in a complete graph of n nodes.

   By employing this formula in our framework, we can efficiently determine all potential dyad combinations from any given set of participants.

4. **FFT Application on Windowed Distance Data:** We iteratively apply

the FFT algorithm to dyadic pairs using the scipy library [143]. After the FFT application, we extract the spectrum's real and imaginary components.

5. **Phase Angles Determination:** Using FFT on distance data yields complex values. We then extract magnitude and phase values using the numpy library [144]. Each participant's data undergoes FFT to determine phase angles for every frequency bin and time step.

6. **Relative Phase Angle Computation between Dyads:** We compute the relative phase angles (differences) for all pairs after determining each participant's phase angles. For our example, this involves computing the differences for each time step and frequency bin.

7. **Windowed Phase Locking Values:** Using relative phase values, we compute PLVs. We align each window element with its counterpart in other windows, resulting in a set of PLVs equal to the window's length.

8. **Dyadic Synchronisation through Averaged Phase Locking Values:** From the derived PLVs, we generate an array equivalent to a single window's length. We can apply a cut-off frequency to remove specific frequencies, omitting the DC component. After calculating the PLVs, we compute the averaged PLV (avgPLV) across pertinent frequency bins, yielding the final dyadic synchronisation value. The computation of PLV and avgPLV follows Equation (3.5).

$$\begin{aligned} PLV_j &= \left| \frac{\sum_{i=0}^{n} e^{i\,\theta_{R(ij)}}}{n} \right| \\ avgPLV &= \frac{\sum_{j=0}^{k} PLV_j}{k} \end{aligned} \tag{3.5}$$

where $i \in \{0..n\}$, $j \in \{0..k\}$ and n are the number of windows, k is the number of relative phase angles in each window, and $\Theta_{R(ij)}$ represents the relative phase angle present in each window i at position j. The value ranges from 0 to 1, where 1 indicates perfect synchrony and 0 no synchrony.

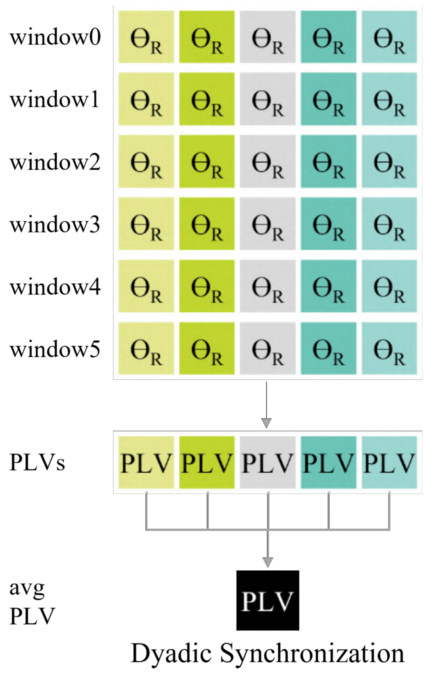


Figure 3.5: Simulated Example illustrating steps 7 and 8.

huSync's potential is demonstrated through this instance, highlighting its capability to process multi-modal data and extract significant conclusions. This tool enables a comprehensive analysis of dyadic synchronisation in small-group settings, proving invaluable for those keen on decoding group dynamics. The results for all pairs of Brahms can be found in Table A.3 and those for Borodin in Table A.4 (see Supplementary Analyses). This practical example reveals our framework's application in musical synchronisation understanding and suggests its applicability in domains where group dynamic comprehension is imperative.

3.3.2 Pseudo-code Representation of huSync's Computational Process

Following the previous section, where we present our computational model, below is a pseudo-code outline of its computational steps, detailing the procedural logic and steps to process input data, extract features, compute synchronisation measures, and validate the results against set criteria.

Algorithm 1 Pseudo-code for Computing Dyadic Synchronisation using huSync Model

```
Function ComputeDyadicSynchronisation(videos)
for all video in videos do
    audio_signal, video_signal ← ExtractSignal(video)
    keypoints_data ← MultiPersonPoseEstimation(video_signal)
    acoustic_features ← ExtractAudioFeatures(audio_signal)
    dyadic_data ← []
    for all person in keypoints_data do
        trajectory ← ExtractTrajectory(person, key='nose')
        dyadic_data.append(trajectory)
    end for
    synchronisation_values ← ComputeSynchronisation(dyadic_data)
    if ValidateResults(synchronisation_values, acoustic_features) then
        StoreResults(synchronisation_values)
    end if
end for
Function ComputeSynchronisation(dyadic_data)
processed_data ← []
for all data in dyadic_data do
    distance ← ComputeEuclideanDistance(data)
    windowed_data ← ApplySlidingWindow(distance)
    fft_values ← ApplyFFT(windowed_data)
    phase_angles ← ExtractPhaseAngles(fft_values)
    relative_phase ← ComputeRelativePhase(phase_angles)
    plv_values ← ComputePLV(relative_phase)
    processed_data.append(plv_values)
end for
avgPLV ← ComputeAvgPLV(processed_data)
return avgPLV
```

3.4 An Instance of huSync to Measure the Directionality of Influence Among Musicians

Our computational model can assess the directionality of influence within pairs of co-performers and examine leadership dynamics during performances. We performed two studies that quantified and analysed the directionality of information flow between participants.

3.4.1 Directionality Assessment in huSync

Within *huSync*, we extend the entrainment block to assess directionality. We use the key-point position data from video frames to compute the directionality of information flow between dyadic pairs of performers using GC. It is a statistical metric that predicts one time series based on another's past values. In this context, time-series data from two performers (designated as X $\rightarrow$ Y) undergo the Granger test to check the predictability of X by Y and vice versa (Y $\rightarrow$ X) - thus helping us compute the directionality of influence between dyads.

3.4.2 Procedure for the Assessment of the Directionality of Influence in Study 2

For study 2 [14], we utilised a subset of the dataset used in Study 1 and included Omega Ensemble's performance of Johannes Brahms' Clarinet Quintet in B minor. We used an experimental approach with binary values to study the directionality of coupling between ensemble musicians, and it consists of four steps:

1. **Input Signal**: Video footages underwent pre-processing using the AlphaPose algorithm, producing detailed trajectory datasets of spatial coordi-

nates.

2. **Feature Distillation**: Among various facial landmarks from AlphaPose, the nose key point was identified as pivotal for tracking head sway (similar to Study 1). This data discerned the directionality of influence among performers.

3. **Measuring Entrainment - Directionality**: The GC technique quantified the flow and directionality of information. Adjustments were made for musical performance delays, examining a range of lag lengths for both X and Y coordinates. Our study used a lag length of 1 second or 30 frames (equal to the video's sampling rate).

4. **Analysis**: The GC analysis outputs were structured for comprehensive review, as demonstrated in Fig. 3.6. Based on previous studies, each matrix cell was binary-coded, indicating Granger causality presence or absence [145; 146]. The GC results for all dyadic pairs were then entered into an ANOVA by selecting appropriate between and within-subject factors. Our study used Head Sway as a within-subjects factor, while Pair and Texture as a between-subjects factor.

	X Coordinate				
	Granger Causing				
	1	2	3	4	5
1	NA	Yes	NC	Yes	Yes
2	Yes	NA	NC	Yes	Yes
3	NC	NC	NA	NC	NC
4	Yes	Yes	NC	NA	Yes
5	Yes	Yes	NC	Yes	NA

→

	X Coordinate				
	Granger Causing				
	1	2	3	4	5
1	NA	1	0	1	1
2	1	NA	0	1	1
3	0	0	NA	0	0
4	1	1	0	NA	1
5	1	1	0	1	NA

(A)

	Y Coordinate				
	Granger Causing				
	1	2	3	4	5
1	NA	Yes	NC	Yes	Yes
2	Yes	NA	NC	Yes	Yes
3	NC	NC	NA	NC	NC
4	Yes	Yes	NC	NA	Yes
5	Yes	Yes	NC	Yes	NA

→

	Y Coordinate				
	Granger Causing				
	1	2	3	4	5
1	NA	1	0	1	1
2	1	NA	0	1	1
3	0	0	NA	0	0
4	1	1	0	NA	1
5	1	1	0	1	NA

(B)

Figure 3.6: Mapping process of GC results to binary values.

3.4.2.1 Pseudo-code for Directionality Assessment in Study 2

Algorithm 2 provides a structured pseudo-code representation of the directionality assessment procedure in Study 2, capturing the computational steps involved in processing video data, applying Granger Causality tests, and statistically evaluating the influence directionality among instrumentalists in musical ensembles using the binary-coded approach.

Algorithm 2 Directionality Assessment Algorithm in huSync Framework for Study 2

```
Function AssessDirectionalityStudy2(videos)
for all video in videos do
    Step 1: Input Signal
    trajectory_data ← ApplyAlphaPose(video)
    Step 2: Feature Distillation
    head_sway_data ← ExtractNoseKeypoint(trajectory_data)
    Step 3: Measuring Entrainment - Directionality
    adjusted_data ← AdjustForPerformanceDelays(head_sway_data)
    gc_results ← []
    for all dyad in GetAllDyads(adjusted_data) do
        lag_data_X, lag_data_Y ← ComputeLags(dyad, lag_length=30)
        gc_result ← ApplyGrangerTest(lag_data_X, lag_data_Y)
        gc_results.append(gc_result)
    end for
    Step 4: Analysis
    binary_coded_matrix ← BinaryCodeGCResults(gc_results)
    anova_results ← ApplyANOVA(binary_coded_matrix, within_factor='Head
    Sway', between_factors=['Pair', 'Texture'])
    if ValidResults(anova_results) then
        StoreResults(anova_results)
    end if
end for
```

3.4.3 Procedure for the Assessment of the Directionality in Study 3

Building on Study 2, in Study 3 [15], we investigated the effects of variations in musical texture on leadership dynamics in musical ensembles using the entire dataset utilised in Study 1. We used GC measures to assess the directionality of interpersonal coupling between instrumentalists. Figure 3.7 depicts the system architecture, modelled as an instance of the computational framework illustrated in Figure 3.1 and an extension of the model illustrated in Figure 3.2. The process proceeds through a systematic sequence of steps, commencing with the selection of videos that align with our analysis criteria as discussed in the next chapter and Tables 4.1 and 4.2. This architecture encompasses the following four integral phases:

1. **Video Frame Extraction**: To ensure stationarity of the data, we apply first-order differencing to the log-transformed time-series data, a standard practice for GC based on previous studies.

2. **Applying the Granger Test**: In this phase, we process video signals using AlphaPose, a multi-person pose estimation algorithm. This processing yields trajectory data, which we save as a JSON file. This data captures the movements of various body parts at a 30Hz sampling rate. We identify the nose key point as the optimal head representation among the various data points. We then use this key point to generate kinematic data, creating position time-series data for subsequent analyses. To ensure the time series remains stationary, we apply first-order differencing to the log-transformed series, following methodologies established in prior research [147; 148].

3. **Directionality Analysis via Granger Causality**: Here, we employ the *grangertest* function from the lmtest package [149] in R Studio to evaluate

the directionality of information flow. The function measures the predictive relationship between two time series. X represents the first set of time series, Y represents the second set, and order represents the number of lags used (default value is 1). We applied this function to the horizontal dimension of the time-series data to account for anterior-posterior head motion and test all possible combinations of dyads. For instance, for performers J and K, we performed tests to examine the directional influence of performer J's head motion on performer K's head motion (X $\rightarrow$ Y) and vice versa (Y $\rightarrow$ X) to determine whether Granger causality exists between the two time series. If the p-value yielded by the Granger test was less than .05 (our criterion for statistical significance), we rejected the null hypothesis and inferred a statistically significant predictive relationship between the two time series in the specified direction. To account for the delay between stimulus and response that is common in musical performances [116; 150], as well as in most behaviours, we performed GC tests for a lag of up to 1 second, setting the order to 30 (equal to the sampling rate) to test for multiple lag-lengths, and examine the nose key-point separately for each pair in both possible test directions (X $\rightarrow$ Y and Y $\rightarrow$ X). We extracted the GC values, including F and p values from each Granger test, and recorded the binary numbers [0,1] indicating the outcome of each test. A value of 1 indicates a statistically significant predictive relationship (assumed to indicate influence or information flow) between the performers in the tested direction, while 0 indicates the absence of a significant predictive relationship in that direction. When GC tests in both directions were statistically significant for a given instrumentalist pair, a 1 was coded for both directions. The main analyses were based on the proportion of significant GC tests among pairs of musicians for each phrase. Specifically, binary values indicating whether

(1) or not (0) each Granger test was significant were averaged across all pairs of performers (separately for each test direction, X → Y and Y → X) for each phrase representing the two textures for each piece. These proportion data were then passed onto the next step to analyse the directionality of influence among co-performers to address our research questions.

4. **Statistical Evaluation and Hypothesis Addressal**: At this stage, we conduct a series of statistical analyses on the GC values to evaluate our hypothesis regarding influence directionality. Table 3.1 displays sample results. Columns with the suffix '_M1_M2' represent the GC test outcomes from Musician 1 to Musician 2. Conversely, columns with the suffix '_M2_M1' represent the GC outcomes from Musician 2 to Musician 1. We recorded results for all possible pairs of instrumentalists for each musical phrase. Note that the aim of these analyses was not to determine whether interaction is taking place at greater than chance levels (given that expert ensembles were intentionally coordinating highly rehearsed performances in a public concert setting), or whether GC can capture leader-follower relations in ensembles (since this has been previously demonstrated in a number of studies). Our question instead concerns the quality of interaction, specifically related to leadership relations, and whether these relations vary across conditions in relative terms. The results for all pairs of Brahms can be found in Table A.5 and those for Borodin in Table A.6 (see Supplementary Analyses). In R, we conducted two primary statistical analyses [151; 152], one using a linear mixed-effects model (LMM) to test for general effects of musical texture on GC values, followed by an Analysis of Variance (ANOVA) to address directionality of influence effects related explicitly to melodic leadership. Musical piece (Brahms Quintet and Borodin Quartet) was included as a random effect in these analyses since we did not have

hypotheses regarding the pieces (they comprised a convenience sample that were on the ensemble's program at the time of data collection), but rather were interested effects that generalize beyond them. Shapiro-Wilk tests indicated that proportion data were not normally distributed in some conditions, even following arcsine transformation. Therefore, we report analyses on untransformed data (additional analyses with arcsin-transformed data yielded similar results). However, given these violations of the normality assumption, we also conducted binomial Generalised Linear Mixed Model (GLMM) analyses on raw binary GC values to check whether equivalent effects are obtained. Obtaining consistent outcomes for the LMM and GLMM analyses would provide evidence for the robustness and reliability of results. Such consistency was observed, and we only report the LMM and ANOVA results in the article (because these tests are standard and facilitate comparison with other studies in the literature). Due to the large number of GC tests run per musical excerpt (to assess exhaustive pairwise relations between instrumentalists), we addressed the issue of potential false positives by correcting for multiple comparisons using the Bonferroni method in supplementary analyses reported in the Appendix (see Binomial Generalised Linear Mixed Model (GLMM) analyses and additional analyses with Bonferroni correction). The results were overall consistent with those reported here.

By quantifying and assessing the directionality of influence, our approach to modifying and using *huSync* as a system and framework offers valuable insights into the complex and subtle dynamics of musical performances, showcasing its versatility with robust results across various contexts.

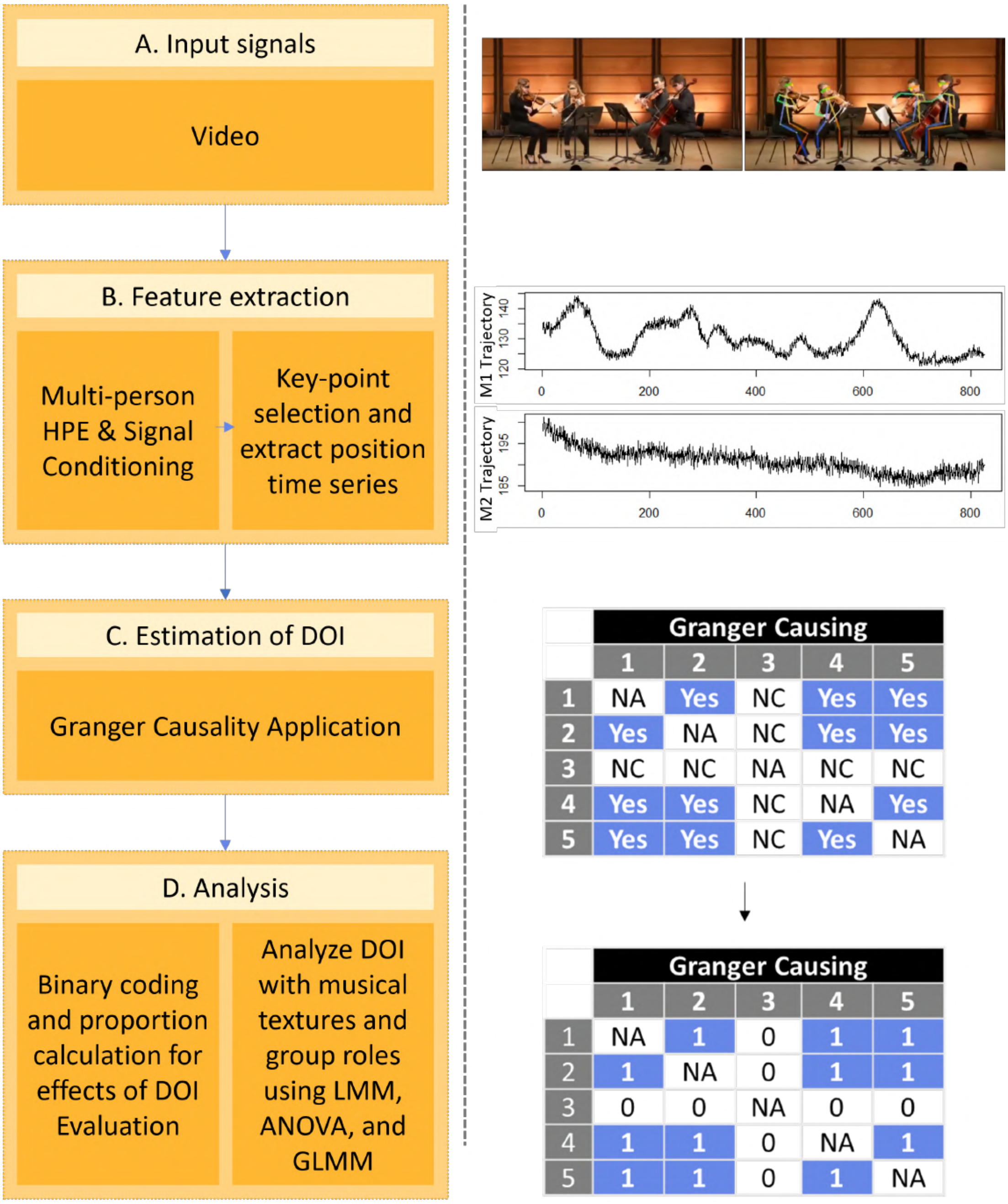


Figure 3.7: An instance and extension of the huSync computational framework and system architecture for assessing directionality of influence. The accompanying images on the right delineate critical phases of the workflow and its intrinsic procedures.

Phrase	M1	M2	T	F M1 >M2	*p* M1 >M2	F M2 >M1	*p* M2 >M1	*p* M1 >M2 (B)	*p* M2 >M1 (B)	M_I
1	m1	m2	P	0.96	0.53	1.38	0.09	0	0	Mixed
10	m1	m2	P	0.92	0.60	1.78	0.006	0	1	Mixed
11	m1	m2	H	1.25	0.17	3.22	0.003	0	1	m5
7	m1	m3	H	0.98	0.49	0.88	0.65	0	0	m5

Table 3.1: This table presents Granger Causality applied to quintet phrases, categorised by texture: Polyphonic (P) or Homophonic (H), and labelled by musician (m1-m5) from left to right as in Fig. 4.1b. The M_I' column labels polyphonic textures as Mixed' reflecting melodic role distribution. The Granger test results for F and p values depict the relationship between musicians (e.g., F M1 >M2 indicates the F value from musician 1 to musician 2).

Table Notes:

- **T (Texture)**: 'P' represents polyphonic textures; 'H' represents homophonic textures.
- **M1 & M2**: Refers to the first and second musician in a dyadic pair.
- **Granger Test Columns**: The headers 'F M1 >M2' and '*p* M1 >M2' (and their counterparts for M2 to M1) show Granger test results for F and p values.
- **(B) Suffix**: Binary values in columns such as '*p* M1 >M2 (B)' signify whether there was a Granger causality between the paired musicians.
- **M_I (Musical Instrument)**: 'Mixed' denotes distributed melodic roles in polyphonic textures.

3.4.3.1 Pseudo-code for Directionality Assessment in Study 3

Algorithm 3 provides a structured pseudo-code representation of the directionality assessment procedure in Study 3, capturing the computational steps involved in processing video data, applying Granger Causality tests, and statistically evaluating the influence directionality among instrumentalists in musical ensembles.

Algorithm 3 Directionality Assessment Algorithm in huSync Framework for Study 3

```
Function AssessDirectionality(videos)
for all video in videos do
    frame_data ← ExtractFrames(video)
    trajectory_data ← ApplyAlphaPose(frame_data)
    stationary_data ← ApplyLogTransformAndDifferencing(trajectory_data)
    gc_values ← []
    for all dyad in GetAllDyads(stationary_data) do
        gc_result ← ApplyGrangerTest(dyad)
        gc_values.append(gc_result)
    end for
    statistical_results ← EvaluateHypothesis(gc_values)
    if ValidResults(statistical_results) then
        StoreResults(statistical_results)
    end if
end for
```

Chapter 4

Dataset and Experiment Design

The empirical study of interpersonal synchronisation and leadership dynamics within small chamber music ensembles necessitates a careful approach to dataset selection and experimental design. This chapter highlights the method to record, collect, and utilize a multimodal dataset tailored to investigate these phenomena. The dataset, comprising high-definition video recordings of concert performances by the Omega Ensemble, a renowned professional chamber music group from Australia, is described in detail. The chapter further elucidates the experimental design that aligns with the research questions being investigated.

The dataset, a critical component of this research, consists of concert performances by the Omega Ensemble. These performances were meticulously selected from multiple locations, focusing on the City Recital Hall in Sydney in 2017. The dataset features compositions by distinguished composers such as Alexander Borodin and Johannes Brahms.

The decision to concentrate on the City Recital Hall recordings was underpinned by technical considerations pertinent to applying a multi-person pose estimation algorithm. The chosen recordings offered an unobstructed view of the performers, thereby minimising occlusions and facilitating the tracking of body

(a) Image from a musical piece composed by Alexander Borodin (1)

(b) Image from a musical piece composed by Johannes Brahms

(c) Image from a musical piece composed by Max Bruch

(d) Image from a musical piece composed by Alexander Borodin (2)

(e) Image from a musical piece composed by Robert Schumann

Figure 4.1: Screenshots of various concert recordings of the Omega Ensemble, illustrating the diversity in the dataset and the rationale for selecting specific recordings for the study.

key points for all ensemble members (See Figures 4.1a and 4.1b). Conversely, the other three video recordings were deemed unsuitable for analysis due to data contamination issues, where extraneous elements such as the audience, microphones, and chairs were visible in the frame (See Figures 4.1c, 4.1d, and 4.1e).

The research specifically focuses on two concert performances: Johannes Brahms' Clarinet Quintet in B minor (Op. 115) and Alexander Borodin's String Quartet No. 1 in A major. These pieces were judiciously chosen based on their structural congruities within the ensemble and the diverse musical textures they manifest, thereby aligning with the research objectives and furnishing a rich context for nuanced analysis.

Details of the recordings are as follows:

1. **Camera and Lens**: Utilisation of Canon 1DX camera body and Canon EF 70-200 1:2.8 L zoom lens, ensuring high-resolution capture.

2. **Video Format**: High-definition QuickTime movies (.MOV) with dimensions of 1920 $\times$ 1080 pixels, 25fps, providing clarity and detail.

3. **Audio Quality**: Recorded in 16-bit stereo at 48 kHz, ensuring fidelity in audio reproduction.

4. **Compression**: Utilisation of the H.264 video codec and Linear PCM audio codec, balancing quality and file size.

5. **Synchronisation**: Achieved using Timecode, ensuring temporal alignment between audio and video streams.

The total duration of the performances was approximately 39 minutes and 13 seconds for the Borodin String Quartet and about 40 minutes and 38 seconds for the Brahms Clarinet Quintet. Specific phrases were carefully selected for analysis, adhering to pre-defined parameters aligning with the research questions, ensuring relevance and applicability. Musicians were labelled for specific positioning as m1, m2, m3, and so on, arranged from left to right (see Fig 4.2).

Figure 4.2: Images from the performance of the Quintet (Top Left) and the Quartet (Bottom Left), along with the outputs available with tracked key points using a pose estimation algorithm (Top Right and Bottom Right). The figure shows the arrangement of musicians from left to right labelled as m1 to m4 (Quartet: m1 - violin1, m2 - violin2, m3 - viola, m4 - cello) and m1 to m5 (Quintet: m1 - violin1, m2 - violin2, m3 - viola, m4 - cello, m5 - clarinet).

4.1 Rationale for Dataset Selection

The dataset was selected with careful consideration of several significant factors that align with the research questions:

1. **Consistent Ensemble Structure**: The Omega Ensemble's consistent structure facilitates a controlled analysis of interpersonal dynamics, directly addressing RQ1 and RQ3 concerning synchronisation and leadership dynamics within the ensemble.

2. **Articulated Musical Textures**: The performances exhibit clearly delineated musical textures, providing distinct cases for the study of synchronisation (RQ2) and directionality of influence (RQ4) in different musical contexts.

3. **High-Quality Recordings**: Utilising high-quality video equipment ensures accurate data for analysis, essential for RQ5's focus on computational methods and the application of advanced algorithms.

4. **Minimal Occlusions and Contamination**: Recordings from the City Recital Hall were specifically chosen due to their minimal occlusions, facilitating the tracking of body keypoints using a multi-person pose estimation algorithm. Other video recordings were limiting, with issues such as audience visibility contaminating the data, thereby restricting their suitability for the study.

The Omega Ensemble provided written consent to use their performance videos, and an ethics committee approved the data collection protocol.

4.2 Annotation and Segmentation of the Dataset

The video recordings were systematically annotated and segmented into sections, each containing a single musical phrase performed by the ensemble. This segmentation was essential for testing the hypothesised effects of musical texture and phrase position on interpersonal synchronisation and directionality of influence.

The annotation process was conducted using ELAN [153], a comprehensive annotation tool for multimedia files. The videos were annotated following a musicological analysis of the performed musical pieces' published score, resulting in annotations that included:

1. **Phrasing**: The division of the music into meaningful phrases, allowing for the study of synchronisation within specific musical contexts.

2. **Textural Classification**: Identification of polyphonic or homophonic textures, facilitating the analysis of directionality of influence in different musical settings.

3. **Instrumentation**: The number of instruments playing at a given time and their roles (e.g., melody, counter melody, harmonic accompaniment), providing insights into the ensemble's structural dynamics.

The annotation process aimed to test the influence of two primary factors on the strength of interpersonal coupling among ensemble performers:

1. **Position within the Musical Phrase**: The start, middle, or end of the phrase, relevant to RQ2 and supporting H3, providing insights into synchronisation patterns.

2. **Musical Texture**: Polyphonic or homophonic, relevant to RQ2 and RQ3, and supporting H2 and H4, allowing for the study of directionality of influence.

The bar numbers from the score corresponding to each excerpt are provided in Table A.1 (For Quintet: Brahms, J. (1892). Clarinet Quintet, Op. 115. N. Simrock.) and Table A.2 (For Quartet: Borodin, A. (1884). Refer Supplementary Analyses). Specific phrases were selected for each recording to adhere to these musical characteristics. Each selected phrase was approximately 30 seconds long and was split into three equal segments to represent the phrase's start, middle, and end. Additionally, it was ensured that all instruments in the ensemble were

Piece	Texture	Minimum	Maximum	Median	Average	Count
Brahms	Homophonic	15.032	38.199	19.742	21.573	27
	Polyphonic	15.488	33.08	23.100	23.528	20
Borodin	Homophonic	15.295	24.973	18.317	19.117	10
	Polyphonic	15.142	29.628	21.271	21.013	11

Table 4.1: Summary of the complete dataset duration and count for our experiments.

Piece	Texture	Minimum	Maximum	Median	Average	Count
Brahms	Homophonic	16.161	30.433	20.615	21.856	12
	Polyphonic	15.488	27.553	20.161	21.105	12
Borodin	Homophonic	15.295	24.973	18.317	19.117	10
	Polyphonic	15.142	29.628	20.924	20.800	10

Table 4.2: Summary of the selected phrases duration and count for Studies 1 and 3.

playing for most of the phrase, allowing for a comprehensive analysis of the ensemble's dynamics.

4.3 Experiment Design

The experiment design was planned to align with the research objectives, focusing on two core objectives: the study of interpersonal synchronisation and the directionality of influence within the ensemble. Both studies leveraged the same dataset to evaluate the proposed computational models.

4.3.1 Study of Synchronisation

The study of synchronisation aimed to quantify the interpersonal coordination within the ensemble, using Phase Locking Values (PLV) computed between head motion trajectories for all possible pairings of musicians.

1. **Quantification of Synchronisation**: PLVs were derived by applying

Piece	Texture	Minimum	Maximum	Median	Average	Count
Brahms	Homophonic	16.75	34.87	21.60	23.51	12
	Polyphonic	15.49	27.55	20.16	21.11	12

Table 4.3: Summary of the duration and count of the selected phrase for Study 2.

FFT to extract phases, calculating phase angle differences between co-performer signals, and averaging the resulting vectors. This produced a dyadic PLV time series reflecting the synchronisation of head sway.

2. **Analysis of Musical Texture and Phrase Position**: ANOVAs were conducted on the PLV outputs to test for main effects of musical texture (homophonic vs. polyphonic) and phrase position (start vs. middle vs. end). Texture addresses coordination under different leadership conditions, while position explores points of heightened structural uncertainty.

3. **Relevance to Research Questions**: This study addresses RQ1, RQ2, RQ3 and RQ5, shedding light on how musical context modulates small ensemble interpersonal dynamics and synchronisation patterns.

4.3.2 Study of Directionality of Influence

The study of directionality of influence sought to assess directional leadership relations within the ensemble, focusing on Granger Causality (GC) on the head motion data.

1. **Assessment of Leadership Relations**: GC uses autoregressive modelling to measure directed predictive influence between two time-series. Significant GC in one direction implies that musician A is driving or leading musician B, providing insights into the leadership dynamics within the ensemble.

2. **Application to Musical Context**: The study was tailored to the musical textures and phrase positions identified in the dataset, allowing for a nuanced analysis of how leadership relations evolve within specific musical contexts.

3. **Relevance to Research Questions**: Studies using GC address RQ2, RQ3, RQ4, and RQ5, exploring the directionality of influence and leadership relations within the ensemble, contributing to the understanding of interactive synchrony in diverse team settings, particularly considering structural elements such as texture.

4.4 Dataset Summary and Phrase Selection

Table 4.1 details the complete dataset's composition and selection. For Studies 1 and 3, the summary is available in Table 4.2, while for Study 2 it is available in Table 4.3. These tables reflect the careful selection process to identify phrases that align with specific study criteria, demonstrating the consideration given to the dataset's relevance to the research objectives. Based on this information, phrases that had consistent texture and predominantly all instruments playing throughout were selected for analysis.

1. **Alignment with Study Criteria**: The selected phrases were chosen with an eye to the study's emphasis on interpersonal synchronisation and directionality of influence among co-performers.

2. **Segmentation of Phrases**: Each phrase, approximately 30 seconds in length, was divided into three equal segments to represent the start, middle, and end of the phrase. This segmentation allowed for a more complete

analysis of the ensemble's dynamics for Study 1 since RQ2 addresses phrase positioning as well.

3. **Comprehensive Analysis**: The analysis of the dataset, as represented in Table 4.1, provides insights into ensemble performance dynamics. The systematic breakdown and annotation of the performances contribute to a deeper understanding of the factors that may influence interpersonal synchronisation and directionality of influence among co-performers.

The careful selection, annotation, and analysis of the dataset and solid experiment design are essential in working towards the research objectives outlined in Section 1.2. The dataset's nature and the attention to detail in its preparation offer a platform for testing the proposed computational models' potential in analysing ensemble music performances, addressing RQ1 to RQ5.

The thoughtful handling of the dataset and the experiment design may contribute to future work in the field of human-computer interaction in music performance analysis. It adds to the understanding of ensemble dynamics and provides a methodological approach that could be useful in future research. This work may also extend insights into nonverbal signalling, entrainment, and emergent leadership in musical groups to small teams.

Chapter 5

Results

This chapter presents results from three studies that collectively use the huSync computational system and model to explore the intricate dynamics of synchronisation and directional influence among ensemble musicians. While the first study focuses on quantifying synchronisation between co-performers, the subsequent two studies broaden the scope to explore the directional influence and the underlying leadership relations within the ensemble. To reiterate and for the sake of convenience, the following are the three primary studies:

1. **Study 1**: S. R. Sabharwal, M. Varlet, M. Breaden, G. Volpe, A. Camurri and P. E. Keller. *huSync - A Model and System for the Measure of Synchronisation in Small Groups: A Case Study on Musical Joint Action*, in IEEE Access, vol. 10, pp. 92357-92372, 2022.

2. **Study 2**: S. R. Sabharwal, Arianna Musso, Matthew Breaden, Eva Riccomagno, Antonio Camurri, Peter E. Keller. *Analysing directionality of influence among ensemble musicians using Granger Causality*, in International Conference of Kansei Engineering and Emotion Research (KEER), Barcelona, Spain, 2022.

3. **Study 3**: S. R. Sabharwal, M. Breaden, G. Volpe, A. Camurri and P. E. Keller. *Leadership Dynamics in Musical Groups: Quantifying Effects of Musical Structure on Directionality of Influence in Concert Performance Videos*. Currently under review.

Drawing from the same dataset, as delineated in Chapter 4, each study leverages high-definition multi-camera video recordings of performances by a professional chamber orchestra. These recordings are carefully annotated based on musical structure, allowing for a nuanced analysis. The Phase Locking Values (PLVs), derived from the musicians' head motion trajectories, serve as a quantitative synchronisation measure. Concurrently, Granger Causality modelling of these head motion trajectories sheds light on the directional influence within the ensemble.

Study 1 explores how musical texture and phrase position influence synchronisation strength between musician pairs. The analysis contrasts textures that feature a distinct melodic leader with those that foster more distributed coleadership. Additionally, it compares synchronisation at points of high structural uncertainty with those of low uncertainty, revealing the subtle interplay between these factors.

Studies 2 and 3 extend the investigation to directed leadership relations, probing the dynamics between musicians under varying musical roles and textures. These studies test hypotheses concerning the strength of causal linkages from melodic parts to accompanying sections, uncovering the complex web of influence within the ensemble.

Together, these studies underscore the sensitivity and applicability of the huSync model in quantifying and analysing the nuanced effects of musical structure on ensemble interpersonal coordination. The results provide fresh insights into the subtleties of nonverbal communication, entrainment, and emergent leadership in musical groups and hold broader relevance for understanding team dynamics in diverse settings. Ultimately, the findings advance the fundamental comprehension of interactive synchrony, enriching both the musical domain and the broader field of joint action.

5.1 Results from Study 1

In Study 1, we conducted our analyses on 44 phrases, summarised in Table 1, carefully selected to meet our criteria. The chosen dataset ensured a good balance between polyphonic and homophonic textures while also considering the duration of each phrase and the quality of the data received on pre-processing videos with a pose estimation algorithm.

Firstly, we present the Phase Locking Values (PLVs) results descriptively, then detail the Analysis of Variance (ANOVA) findings. We analysed the performances of the Brahms and Borodin pieces individually due to the differing number of performers in each piece. For each piece, we entered the PLVs of all pairs into an ANOVA analysis. This analysis included Phrase Position (Start, Middle, End) as a within-subjects factor and Texture (Homophonic, Polyphonic) and Pair (referring to each separate pairing of individuals from the ensemble) as between-subjects factors. We included the factor 'Pair' in the analyses because no two instrumentalists played identical parts, and the specific pairing of parts might systematically influence PLV.

Nevertheless, we did not delve into a detailed analysis of such potential effects as it was outside the scope of our study. We incorporated phrase duration as a covariate in the analyses to account for its potential impact on PLV. We conducted the ANOVAs using jamovi software. Fig. 5.1 graphically depicts the PLV results for the three phrase positions in the two textures. We averaged the data across pairs for Brahms and Borodin performances. These graphs demonstrate that polyphonic textures typically exhibit higher PLVs than homophonic textures, aligning with our hypothesis outlined in section 1.2. We also noted that PLVs start at a lower value in both textures. In the polyphonic texture, values commence relatively high, rise in the middle of the phrase, and then decline towards the phrase's end. Conversely, in the homophonic texture, values start lower and maintain this level until a slight increase at the phrase endings.

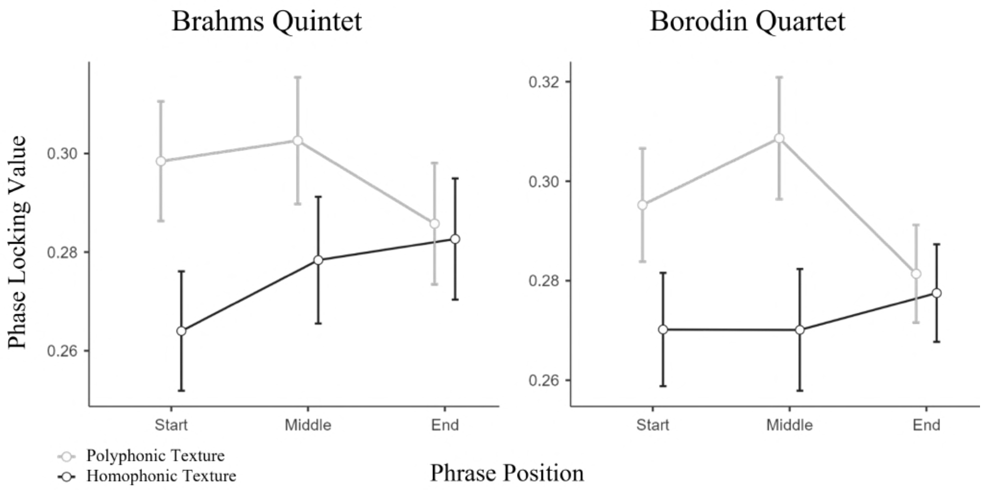


Figure 5.1: Phase locking values, indicating synchronisation of co-performers head motion, across phrase positions for polyphonic and homophonic texture for Position x Texture in the Brahms and Borodin pieces).

In Fig. 5.2, results are shared as network plots of averaged PLVs across all phrases for individual instrument pairings observed for the Brahms and Borodin performances separately. Here, most pairs show a higher level of synchronisation in polyphonic textures at the start, middle, and end of each phrase, suggesting that the effect is general and not tied to specific instrument pairings.

The ANOVA results are illustrated for the Brahms performance in Table 5.1 and for Borodin in Table 5.2. Values highlighted in bold indicate statistical significance ($p < 0.05$). For Brahms, the ANOVA revealed a statistically significant

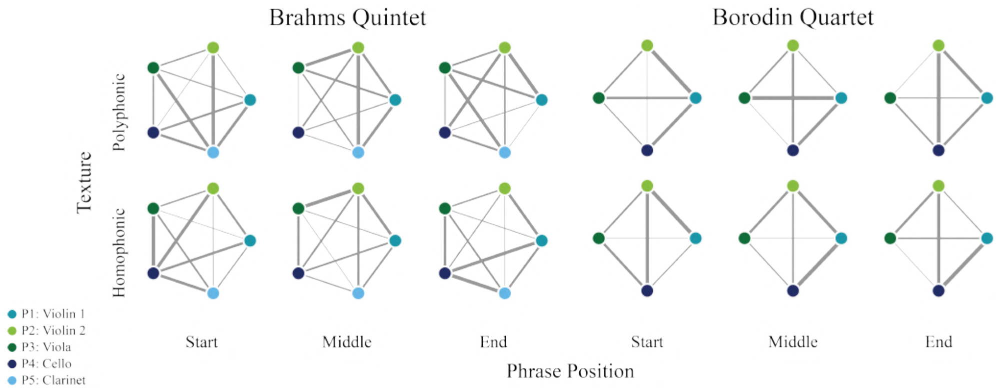


Figure 5.2: Network plots for ensemble PLV data by instrument for each condition (texture and phrase position) in Brahms and Borodin pieces. Edge thickness indicates the coupling strength based on phase locking values averaged across all phrases. Each coloured node indicates an instrument played by the performers.

main effect of Texture, $F(1, 219) = 16.08$, $p < 0.001$, and a significant two-way interaction between Position and Texture, $F(2, 438) = 6.098$, $p = 0.002$. For Borodin, there was also a statistically significant main effect of Texture, $F(1, 107) = 14.051$, $p < 0.001$, and a significant two-way interaction between Position and Texture, $F(2, 214) = 3.399$, $p = 0.035$. For both Brahms and Borodin, the main effect of position was not statistically significant.

Between Subjects Effects					
	Sum of Squares	**df**	**Mean Square**	**F**	**p**
Pair	0.0126	9	0.0014	0.25	0.986
Texture	0.0902	1	0.09024	16.081	**<.001**
Pair ∗ Texture	0.0583	9	0.00648	1.155	0.326
Duration	0.2595	1	0.25948	46.244	<.001
Residual	1.2289	219	0.00561		

Within Subjects Effects					
	Sum of Squares	**df**	**Mean Square**	**F**	**p**
Position	0.00598	2	0.00299	1.003	0.368
Position ∗ Pair	0.08754	18	0.00486	1.632	0.049
Position ∗ Texture	0.03634	2	0.01817	6.098	**0.002**
Position ∗ Duration	0.00424	2	0.00212	0.711	0.492
Position ∗ Pair ∗ Texture	0.03341	18	0.00186	0.623	0.882
Residual	1.30509	438	0.00298		

Table 5.1: ANOVA results for Between and Within Subjects Effects for the Brahms Concert.

Between Subjects Effects					
	Sum of Squares	**df**	**Mean Square**	**F**	**p**
Pair	0.02003	5	0.00401	1.542	0.183
Texture	0.0365	1	0.0365	14.051	**<.001**
Pair ∗ Texture	0.00412	5	0.000825	0.318	0.901
Duration	0.24612	1	0.24612	94.746	<.001
Residual	0.27796	107	0.0026		

Within Subjects Effects					
	Sum of Squares	**df**	**Mean Square**	**F**	**p**
Position	0.01229	2	0.00614	2.849	0.06
Position ∗ Pair	0.01569	10	0.00157	0.727	0.698
Position ∗ Texture	0.01468	2	0.00734	3.403	**0.035**
Position ∗ Duration	0.00972	2	0.00486	2.253	0.108
Position ∗ Pair ∗ Texture	0.0302	10	0.00302	1.4	0.182
Residual	0.46151	214	0.00216		

Table 5.2: ANOVA results for Between and Within Subjects Effects for the Borodin Concert.

Overall, these results indicate that for both pieces, PLVs were reliably higher — hence, interpersonal coupling between performers was stronger—for polyphonic than homophonic textures. Nevertheless, this effect of texture varied throughout musical phrases. Specifically, the end of phrases diminished the texture effect because coupling strength decreased in polyphonic textures and increased in homophonic textures.

5.1.1 Analysis of audio features

While our principal analysis focuses on ensemble coordination of co-performer body motion, we conducted an additional analysis to examine the relationship between the synchronisation of body movements, which provides visual cues, and ensemble sounds.

		Texture					
		Homophonic			**Polyphonic**		
		Phrase Position					
Piece	**Measure**	**Start**	**Middle**	**End**	**Start**	**Middle**	**End**
	Pulse Clarity						
	Mean	0.111	0.122	0.167	0.148	0.126	0.135
Brahms	SD	0.046	0.057	0.058	0.075	0.047	0.052
	Event Density						
	Mean	1.416	2.098	1.767	1.733	2.105	2.205
	SD	0.844	1.484	0.931	0.846	0.819	0.845
	Pulse Clarity						
	Mean	0.153	0.157	0.152	0.16	0.156	0.144
Borodin	SD	0.063	0.044	0.083	0.081	0.054	0.058
	Event Density						
	Mean	2.102	2.202	1.631	2.197	2.025	2.033
	SD	0.946	0.934	0.799	0.798	0.792	1.25

Table 5.3: Mean and standard deviation (SD) of estimates of pulse clarity and event density as a function of texture (homophonic and polyphonic) and phrase position (start, middle, and end) for performances of pieces by Brahms and Borodin.

Because we do not have multi-track audio recordings for each instrument on a separate track, we computed indirect measures of global ensemble synchronisation from stereo auditory recordings of the entire ensemble sound. Drawing from prior research [154; 155; 156], we incorporated estimates for 'pulse clarity' and 'event density'. We calculated these estimates using the 'mirpulseclarity' and 'mireventdensity' functions from the MIRtoolbox in MATLAB [157]. 'Pulse clarity' reflects the strength of rhythmic beats, while 'event density' measures the average frequency or number of detected events per second. We present descriptive statistics for these measures in Table 5.3. To assess potential effects related to these audio features, we ran a linear mixed-effects model analysis using the lmer package [158] in R [159] with PLV as the dependent variable, pulse clarity, event density, texture, and phrase position as predictor fixed effects, and piece as a random effect (with intercepts allowed to vary). Pulse clarity values were arcsine-transformed, and event density values were log-transformed before analysis. The results revealed a link between PLV and event density. Expressly, a likelihood-ratio test indicated that a model including event density provided a better fit for the data than a model without it (χ^2 (1)= 7.44, p = 0.006), whereas pulse clarity did not contribute significantly to the model (χ^2 (1) = 0.03, p = 0.884). Examination of the output for the entire model indicated that PLV values increased with increasing event density (β = 0.031, SE = 0.011, t = 2.767, p = 0.006). These results are consistent with a growing body of evidence that visual and audio cues are both relevant in assessing interpersonal synchronisation in musical ensembles [154][155][160][156][133]. In future research using multi-track audio, we can explore the relationship between auditory and visual information more thoroughly and assess the correspondence between leader-follower relations across modalities.

5.2 Results from Study 2

In Study 2, we performed a two-way ANOVA to assess Granger Causality (GC) values. Our statistical analysis used version 1.6.23 of jamovi software (The Jamovi Project, 2021). Based on the estimated marginal means displayed in Fig. 5.3, our results indicate that homophonic textures with clear melodic leadership have higher mean GC values than polyphonic textures with more evenly distributed leadership dynamics. The figure provides a clear visual depiction of the differences in leadership dynamics between the two textures.

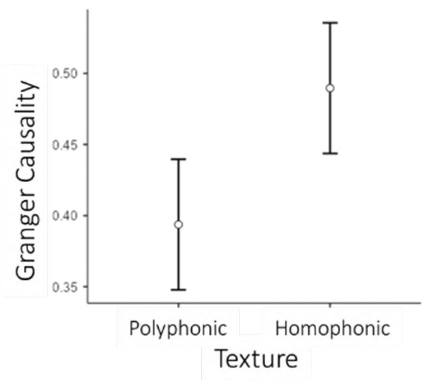


Figure 5.3: A plot showing the mean values of GC for homophonic and polyphonic textures.

We analysed our ANOVA using Luepsen's binary value approach [161] to examine the GC results for all potential dyadic pairs. We considered the Nose trajectory data in X and Y coordinates representing Head Sway as a within-subjects factor. Additionally, we treated Texture (either Homophonic or Polyphonic) and Pair (with each pair represented as M1-M2, where M1 influences M2) as between-subjects factors.

In Table 5.4, we present the ANOVA outcomes for the Brahms concert, distinguishing between the effects observed within and between subjects. To simplify interpretation, we have highlighted values achieving statistical significance in green. The table displays the main effects in three areas: Texture, with F(1,440) = 8.421, p = 0.004; Pair, with F(19,440) = 2.041, p = 0.006; and Head Sway, with F(1,440) = 19.047, p¡.001. Moreover, we observed an interaction between Head Sway and Texture, as evidenced by F(1,440) = 7.928, p=0.005.

Between Subjects Effects					
	Sum of Squares	**df**	**Mean Square**	**F**	**p**
Pair	10.15	19	0.534	2.041	0.006
Texture	2.2	1	2.204	8.421	0.004
Pair ∗ Texture	3.21	19	0.169	0.646	0.871
Residual	115.17	440	0.262		

Within Subjects Effects					
	Sum of Squares	**df**	**Mean Square**	**F**	**p**
Head Sway	4	1	4.004	19.047	$< .001$
Head Sway ∗ Pair	4.16	19	0.219	1.042	0.410
Head Sway ∗ Texture	1.67	1	1.667	7.928	0.005
Head Sway ∗ Pair ∗ Texture	3.67	19	0.193	0.918	0.561
Residual	92.5	440	0.21		

Table 5.4: ANOVA results for Between and Within Subjects Effects for the Brahms Concert.

It is worth noting that the inclusion of Head Sway in the ANOVA was to account for variance related to movement direction. Nevertheless, this variable was not the central focus from a theoretical perspective, so results for its dimensions were not presented.

These initial findings suggested a difference in directional coupling between homophonic and polyphonic textures, providing insights into ensemble performance dynamics. Building on these results, we extended the experiment to perform the study of directionality of influence on the entire dataset, akin to the one for Study 1.

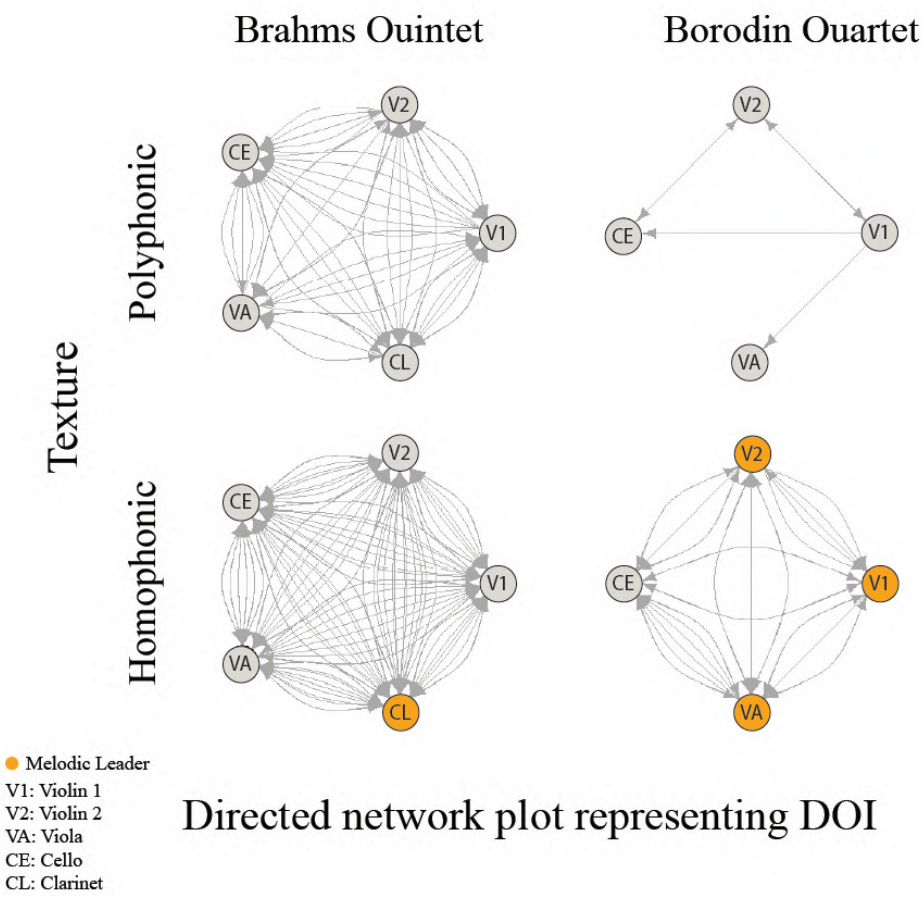


Figure 5.4: Directed network plots for ensemble GC values by instrument for each texture in Quintet and Quartet musical pieces, representing the Directionality of Influence (DOI). Edge direction indicates DOI across all phrases. Homophonic textures have a clear melodic leader, while leadership is assumed to be distributed in polyphonic textures. Each node represents an individual instrumentalist, and the yellow dot indicates the melody instruments (which varied across the analysed phrases for the Quartet) in homophonic textures.

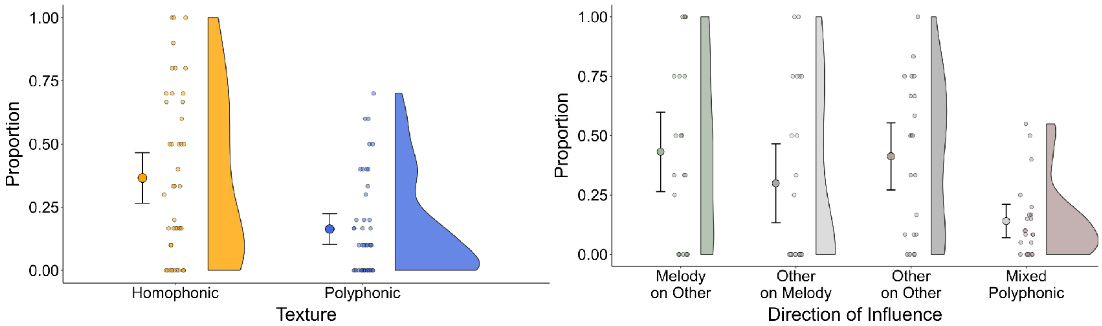


Figure 5.5: (A – Left) Proportion of significant GC values for homophonic and polyphonic textures, averaged across the Quintet and Quartet; (B – Right) Proportion of significant GC values for four categories of musical roles adopted by co-performers indicating the directionality of influence in the musical ensemble.

5.3 Results from Study 3

Study 3 explored the dynamics of interpersonal coupling in musical ensembles, focusing on the directional influence among instrumentalists based on musical textures.

Fig 5.4 shows network plots of GC values representing directional linkages statistically significant for individual instrument pairings as a function of musical texture in the Quintet and Quartet performances. Visual inspection of the plots reveals that connections are denser for the Quintet than the Quartet. This effect was not of interest in the present study, since musical piece was considered as a random effect in our analyses. Note that comparing the pieces would be problematic and inconclusive because they vary on multiple confounding parameters (e.g., number of players, composer, tempi, key, and stylistic elements). Of note, it is also evident that there is a greater density of connections for homophonic than polyphonic textures, consistent with our main hypothesis that interpersonal coupling would generally exhibit higher directionality in homophonic textures with a clear melodic leader compared to polyphonic textures with distributed or changing leadership roles. Evidence for or against our additional specific hypothesis that the melody player would influence other players more than vice versa in homophonic textures is less readily discernable from visual inspection of these network plots alone since numerous connections exist not just between the melody player and other instrumentalists, but also among these other players themselves. This underscores the fact that our observations are grounded in a probabilistic framework, reflecting tendencies not certainties. For example, while findings suggest a melody player often assumes a leadership role, it does not rule out influential interactions from other ensemble members. Separate statistical analyses on the GC values were conducted to address each of the two hypotheses.

The hypothesis that interpersonal coupling would exhibit higher directionality in homophonic textures than polyphonic textures was tested by examining the proportion of instrumentalist pairs exhibiting statistically significant GC test values as a function of musical texture. The results in Fig 5.5(A) revealed a higher proportion of significant GC values for homophonic compared to polyphonic textures, thus confirming the hypothesis. To further evaluate this effect, we computed a LMM using the lme4 package in R [162], with texture as a fixed factor and piece, part, phrase, and direction of the GC test (within each pair of instrumentalists) as random effects. We included the piece as a random effect because our hypotheses were not specific to the two particular musical works featured but rather representative of Western chamber music in general.

A likelihood-ratio test indicated that the full model with texture provided a better fit to the data than a reduced model, including only the random effects ($\chi^2(1) = 10.90$, $p < .001$; Log Likelihood = 17.20 (full) vs 11.80 (reduced), AIC = -20.40 vs -11.50, BIC = -3.10 vs 3.37). The full model revealed a statistically significant effect of texture on GC values (Effect Estimate = -0.248, SE = 0.070, t = -3.54, p = 0.001, 95% CI [-0.388, -0.107]), indicating that the directionality of coupling was reliably higher in homophonic than polyphonic textures.

The second analysis tested the hypothesis that the instrument playing the melody would exert greater influence on instrumentalists playing accompaniment material more than vice versa in homophonic textures. For this analysis, we classified the GC test outcomes based on the musical roles of the instrumentalists, which included melody, accompanying (other), or mixed (in polyphonic settings). We identified four categories of direction of influence, which can be seen in Fig 5.5(B): (1) melody instrument influencing accompanying instruments (Melody on Other), (2) accompanying instruments influencing the melody instrument (Other on Melody), (3) accompanying instruments influencing other accompanying instruments (Other on Other) in homophonic textures, and (4) mixed roles in polyphonic textures. GC values were entered into a LMM with the direction of influence category as a fixed factor and piece, part, and phrase as random effects. This full model provided a better fit to the data than a reduced model with only random effects, as indicated by a likelihood-ratio test ($\chi^2(3) = 15.50, p < .01$; Log Likelihood = -1.06 (full) vs -8.82 (reduced), AIC = 18.10 vs 27.60, BIC = 37.90 vs 40.00).

We conducted a follow-up ANOVA with three planned orthogonal contrasts to determine the specific effects of the direction of influence. These contrasts compared homophonic leadership categories combined (Melody on Other, Other on Melody, and Other on Other) versus the mixed polyphonic category (which is equivalent to the analysis reported above, but with different degrees of freedom), homophonic categories including a melody player (Melody on Other and Other on Melody) versus the homophonic category without a melody player (Other on Other), and the category reflecting melody instrument influence on other instruments (Melody on Other) versus accompanying instrument influence on the melody instrument (Other on Melody). This analysis indicated that GC values were significantly higher for homophonic than polyphonic textures

$t(45.1) = 2.849, p = .0066$, 95% CI [0.212, 1.235] and for melody instrument influence on others than for other instrument influence on the melody instrument $t(46.8) = 2.520, p = .015$, 95% CI [0.027, 0.238]. This latter finding is consistent with our specific hypothesis about musical role. In addition, we found that GC values for homophonic pairings including a melody player were not significantly different from values for homophonic pairings without a melody player $t(46.8) = -1.039, p = .304$, 95% CI [-0.278, 0.089]. This post-hoc finding was not hypothesised but is informative to the extent that it indicates high directionality of influence among accompanying instrumentalists.

5.3.1 Binomial Generalised Linear Mixed Model (GLMM) analyses and additional analyses with Bonferroni correction.

Section 5.3 reports analyses using linear mixed-effects models (LMMs) to test for global effects of musical texture on Granger Causality (GC) values, followed by Analysis of Variance (ANOVA) to address directionality-of-influence effects related to melodic leadership. However, given that the raw data consist of binary GC values (0, 1) that are not normally distributed, we also conducted binomial Generalised Linear Mixed Model (GLMM) analyses on these values to check whether equivalent effects are obtained. In this section, we report these GLMM analyses, which were run using the 'glmer' function from the lme4 (version 1.1-31) package in R (version 4.2.2) within RStudio (version 2022.12.0+353). A separate GLMM analysis addressed each of our two hypotheses. The first GLMM analysis tested the hypothesis that there would be higher directionality in interpersonal coupling in homophonic textures (with a clear melodic leader) than in polyphonic textures (with distributed/changing leadership roles). The GLMM included texture as a fixed factor and piece, part, phrase, direction of the GC test (within each pair of instrumentalists), and instrumentalist pair (coded as unique combinations of instrumentalist numbers) as random effects. Note that instrumentalist pair is included in the GLMM but not in the LMMs in the main article because the binary values were averaged across instrumentalist pairs to get the proportion measure in the latter. A likelihood-ratio test indicated that this full model provided a better fit to the data than a reduced model that included only the random effects ($\chi^2(1) = 8.21, p < .01$; Log Likelihood = -344 (full) vs -348(reduced), AIC = 702 vs 708, BIC = 734 vs 736). For the full GLMM, there was a statistically significant effect of texture on GC values (Effect Estimate = -1.510, SE = 0.499, z = -3.03, p = 0.0025, 95% CI [-2.578 -0.509]). This outcome confirms the LMM result reported in the main article.

The second GLMM analysis examined the effects of the musical roles of the instrumentalists (melody vs accompaniment) on the directionality of coupling. As in the corresponding LMM analysis in the main article, we included four categories of direction of influence: (1) melody instrument influencing accompanying instruments (Melody on Other), (2) accompanying instruments influencing the melody instrument (Other on Melody), (3) accompanying instruments influencing other accompanying instruments (Other on Other) in homophonic textures, and (4) mixed roles in polyphonic textures. The GLMM included direction-of-influence category as a fixed factor and piece, part, phrase, direction of the GC test, and instrumentalist pair as random effects. A likelihood-ratio test indicated that this full model provided a better fit to the data than the reduced model with only random effects (($\chi^2(3) = 15.70, p < .01$; Log Likelihood = -340 (full) vs -348 (reduced), AIC = 698 vs 708, BIC = 740 vs 736). There was a statistically significant effect of direction-of-influence category on GC values in the full GLMM (Effect Estimate = -1.510, SE = 0.499, z = -3.03, p = 0.0025, 95% CI

[-2.578 -0.509]). Again, the outcome is consistent with the LMM result in the main article.

The direction-of-influence effect was broken down using planned orthogonal contrasts. The outcomes were essentially the same as the corresponding analysis on proportional data in the main article. Raw binary GC values were significantly higher for homophonic than polyphonic textures (z = 3.310, p = .0009, 95% CI [1.957, 7.635]) and for melody instrument influence on others than for other instrument influence on the melody instrument (z = 1.972, p = .049, 95% CI [0.005, 1.642]). However, GC values for homophonic pairings including a melody player were not significantly different from values for homophonic pairings without a melody player (z = -1.033, p = .301, 95% CI [-1.665, 0.515]). In sum, the outcomes of the GLMM analyses reported here are in all respects equivalent to the LMM results in the main article.

From the results of our research, we discern that in chamber music ensembles, the nature of the musical texture—whether homophonic or polyphonic—plays a pivotal role in shaping interactions among musicians. In contexts where the music is homophonic, the individual presenting the melody frequently assumes a central guiding role. This deepens our comprehension of the subtle interplay and leadership dynamics at play in group musical performances.

Chapter 6

Beyond core objectives

During our research, we uncovered related topics that expanded on our primary objectives and carried out two secondary projects that built upon our foundational theories, showcasing the versatility of our insights in HCI research. One project offers practical application, while another explores new research challenges. These works demonstrate the applicability and relevance of our primary research in varied contexts of HCI, establishing a link between our principal findings and their extended applications.

6.1 MULTIPLATAGE's ProHome: A Computational Approach to Movement Analysis in Geriatric Healthcare

The ProHome initiative is part of the larger MULTIPLATAGE project (NET-2016-02361805). Its objective is to create a multi-component intervention platform to improve the management of older people with multiple health conditions and medications. proHome is being developed in collaboration with Ospedale

Galliera in Genova, Italy and aims to predict adverse outcomes in older patients. The Ministry of Health, through the Call Finalised Research in 2016, co-funded the MULTIPLATAGE partnership. The partnership comprises healthcare institutions such as EO Ospedali Galliera and Ospedale Policlinico San Martino in Genova, University Federico II of Naples, AOU Maggiore della Carità in Novara, AOU Mater Domini in Catanzaro, ASP Crotone, and Istituto S.Anna in Crotone. These partners are shaping this project, and the insights obtained thus far are playing a pivotal role in ongoing efforts in PNRR, showcasing the extension and applicability of our research.

The project hinges on the Multidimensional Prognostic Index (MPI)[163], a tool with proven accuracy in predicting short and long-term mortality in hospital settings and among older populations. The MPI, unique to geriatric medicine, relies on a comprehensive geriatric assessment (CGA) that evaluates health, functionality, cognitive ability, nutrition, and social aspects using standardised and validated rating scales. It predicts outcomes like mortality, (re) hospitalisation, and institutionalisation.

ProHome employs state-of-the-art sensor technology within a hospital setting designed to mimic a home, targeting older patients from intensive care. The objective is the analysis of biophysical and motor data, and before discharge, these patients undergo a 3-5 day observation to measure their MPI and improve their health metrics. The initial system uses a Fitbit smartwatch and an Azure Kinect depth-sensing camera. The Azure Kinect DK, equipped with AI sensors, offers computer vision and speech models. It integrates a depth sensor, microphone array, video camera, and orientation sensor in one device, supported by various software development kits (SDKs).

This simulated home setup comprises a living room, bedroom, restroom, and a 5-meter corridor. In this method, professionals non-invasively monitor patients to enhance remote monitoring capabilities. The camera in the corridor records the patients' mobility. This data assists in tracking the entire body using the Body Tracker SDK, capturing key body points. This data feeds into the development of biomechanic models, analysing features like body axis variations, movement fluidity, and smoothness, among others.

We developed a software framework to acquire patient data from a Fitbit smartwatch and Kinect. Our framework follows a three-pronged approach: 1) Data Acquisition, 2) Data processing, and 3) Analysis. An ETL pipeline acquires patient data in real time. The Fitbit stores data on the Fitbit cloud, while the hospital server holds Kinect data. The second step, data processing, is where the efforts are rather extensive since although the information for both devices is available in the JSON, the structures differ. We track the heart rate, steps, and sleep from the Fitbit. In contrast, from the Kinect, we developed computational models to track, and compute from, skeletal information to provide consolidated metrics concerning the number of times walked in the corridor, walking speed (max and average), body sway (mean and trend), and anterior-posterior leaning (mean and trend). Our team is now investigating these data points further according to the MPI protocol and we are performing statistical analyses to assess the overall efficiency of this framework for medical care centres.

As of now, we have collected data from over 50 patients and are analysing this information. Our preliminary results show promise, and we are directing our ongoing efforts towards synthesising these observations. We will compile our findings into a research paper, bridging engineering and medical perspectives. This work aims to serve as a blueprint for other medical care centres, guiding them in enhancing care for elderly patients. Further research is anticipated as part of the PNRR Project Spoke 2.

6.2 Examining the Correlation Between Dance and Electroacoustic Music Phrases: A Pilot Study

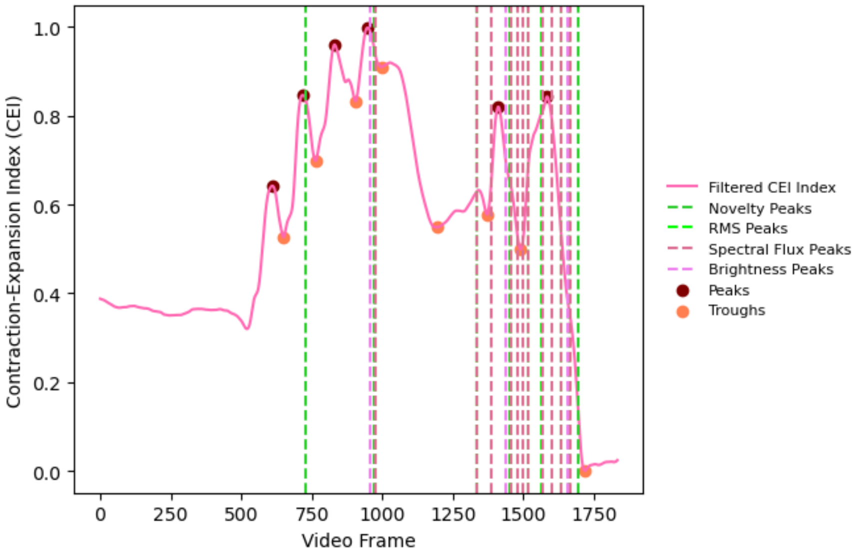


Figure 6.1: Plot indicating CEI with its peaks and troughs, and vertical lines for RMS, Novelty, Spectral Flux, and Brightness peaks for Phrase 3.

In our primary research, we anchored our approach on the flexible *huSync* framework and used musical ensemble performances as a test bed. We employed HPE algorithms to extract kinematic data from video sequences. Initially, we concentrated on fundamental movements like head sway. Recognising a gap, we expanded our scope to understand individual movements and their influencing factors. While HPE algorithms benefit non-invasive studies in natural settings, they can yield noisy data compared to MoCap systems.

Based on the learnings of our primary research studies, we identified potential areas for future work, including developing computational models for movement saliency and relevance. It is crucial to work with reliable, low-noise data for these computational models and to go beyond small-group interactions.

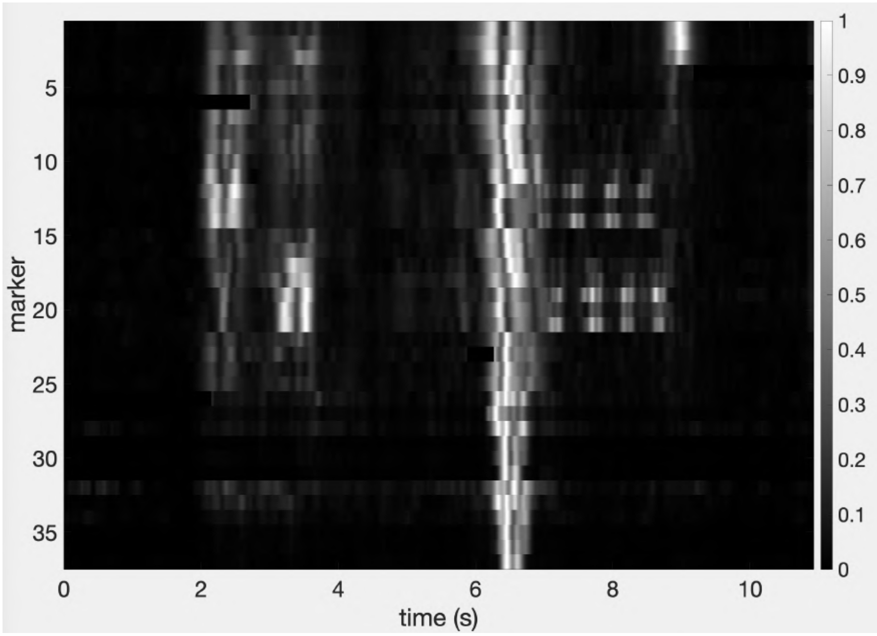


Figure 6.2: Mocapgram showing norm acceleration (m/s) for markers, normalised to 0-1 and shown as greyscale: Head 1-3; upper torso 4-9, 15-16; right arm 10-14; left arm 17-21, lower torso 22-25, right leg 26-31, left leg 32-37.

With this in mind, we conducted a pilot experiment, studying the synchronisation between electroacoustic dance and music phrases. We emphasised the connection between musical accents and spontaneous choreography. Two performers danced to electroacoustic tracks rich in accents but lacking clear rhythm, with their movements captured through a Motion Capture system. The harmony between music and dance is essential in this domain. Our goal was to understand the extent of synchronicity between dance gestures and musical accents, offering a precise, measurable analysis of the auditory-kinematic connection. We utilised audio attributes derived from diverse tracks via the MIRtoolbox, a sophisticated tool designed for music feature extraction. Core kinematic attributes (such as velocity, acceleration, and jerk) and their respective vector norms were derived from refined mocap datasets using Savitzky-Golay filters. To visually represent all these features, mocapgrams were used, providing an insightful qualitative view (refer to Fig. 6.2)[164; 165]. After thorough qualitative scrutiny, we isolated kinematic attributes for seven anatomical segments and evaluated the means of their vector norms. Informed by prior research [166] and the initial visual interpretation of dancer motions in the mocapgram, our early focus gravitated towards the contraction-expansion index (CEI) as a kinematic movement feature.

Central to the study was the temporal binding window (TBW) concept. TBWs define the span within which we perceive auditory and visual stimuli as simultaneous. Based on earlier findings, we adopted an average adult TBW of about 166 ms (equivalent to 20 mocap frames) for brief stimuli like beeps and flashes [167]. Peak detection was carefully tuned using specific algorithms for accuracy and prominence in audio and movement features [168; 169].

A quantitative measure, the 'Jaccard Similarity Index', was employed to determine the degree of overlap between auditory and kinematic features [170]. Synchronicity standards were defined to ensure peaks fall within the TBW. Our preliminary analysis explored the audio-dance dynamic by comparing peak sequences, revealing synchronisation levels across distinct phrases, hinting at the dancers' unique tendencies in aligning movements to varied audio cues.

Results revealed that temporal synchronisation existed between the music's accents and the dancers' movements, extending even to non-rhythmic portions of the music. The study demonstrated that peak synchronisation percentages varied significantly across music and dance phrases. Specific features like 'Novelty' and 'RMS' showed higher alignment in select scenarios, while 'Brightness' had lesser synchronisation. This variance implied underlying patterns and relationships governing these features 6.1. Moreover, the research discerned that the differences between the peak kinematic features for different body parts were extensive for some phrases, aligning with observed distinctions from qualitative evaluations.

We also observed significant variation, sometimes up to 50%, in kinematic feature peaks across different body parts for some phrases. Median values for acceleration peaks ranged between 4.55 and 26, reflecting diverse synchronicity patterns among dancers. This varied alignment aids in understanding the underlying interplay between music and dance.

Our findings suggest that the dynamic relationship between audio features and movement patterns in dance significantly depends on individual preferences in choreography. The analysis underscores the strong link between auditory stimuli and human kinematics, essential for fields like dance pedagogy and performance analysis. As mentioned earlier, the findings lay the groundwork for future research by helping us narrow down the features to study saliency in movement and music. Immediate next steps will involve using more sophisticated statistical tests exploring other movement and audio features that help investigate saliency in music and dance [45; 171], and an elaborate validation studying with up to 100 participants.

Chapter 7

Discussion

The objective of this thesis was two-fold. Firstly, it aimed to create a computational framework and system called *huSync*, which allows for analysing non-verbal social-communicative behaviour in small-group interactions. In the intersecting fields of software engineering and interaction design, *huSync* is a methodological contribution to HCI, particularly emphasising embodied interaction design and somatics. These specialisations are integral to the extensive research efforts of the InfoMus-Casa Paganini team.

Researchers can leverage *huSync* on video sequences to conduct studies in naturalistic environments without the usual interference associated with motion capture setups. Our developed solution aspires to motivate designs that centre on the human body and collective tasks, providing insights and tools where the human body is the focal point of new interaction projects [172; 173; 174].

Secondly, *huSync* was tested in scenarios investigating the correlation between interpersonal coordination and the directionality of influence among co-performers and musical structures. In such scenarios, it emerges as a pragmatic alternative to traditional systems, such as MoCap suits, for measuring dyadic synchronisation and the directionality of influence between co-performers in musical ensembles through the automated analysis of human body motions. The outcome of these investigations is both methodological and empirical, providing insights into technical aspects and conceptual issues relevant to the study of real-time human interaction and non-verbal communication in naturalistic settings.

From a methodological perspective, our structured process involves collecting kinematic data from regular video recordings in a non-intrusive and marker-free way. We then analyse this data to determine the synchronisation and directional influence between musicians performing together. We measure this for all possible pairs within the group using phase-locking values and Granger Causality. This approach allows us to obtain information about the interaction between specific individuals, which is not possible with a global and general measure. However, interpreting natural behaviour data can be complex and challenging, unlike controlled experiments where we systematically manipulate independent variables.

As an empirical outcome, we applied the above techniques for body motion analysis to investigate the effects of musical structure in three distinct studies. Study 1 shows two aspects of musical structure—texture and phrase position—on the strength of interpersonal coupling in instrumental ensembles. We then conducted two studies (Study 2 and 3), where we operationalised and extended the *huSync* model to study the directionality of influence. We analyse how head sway influences unfolding leadership dynamics among co-performers.

Our first study found that polyphonic textures showed a higher degree of coupling than homophonic textures. Such increased coupling might occur because performers distribute coupling evenly among themselves in leadership ambiguity in polyphonic textures. In contrast, in homophonic textures, accompanying performers are more closely linked to a single performer who serves as the melodic leader. Previous research on interpersonal coordination in controlled laboratory tasks [70; 175; 176] supports this interpretation. This research suggests that distributed leadership in polyphonic textures could lead to improved ensemble synchronisation due to heightened mutual adaptation, anticipation, and joint attention [2; 29; 35].

While the positioning within musical phrases did not universally influence the intensity of interpersonal synchronisation, the interplay between phrase position and musical texture was apparent. This interaction suggests a more pronounced role of a distinct melodic leader in the initial segments of phrases compared to their latter portions. Notably, we observe enhanced synchronisation in polyphonic structures at the beginning and central parts of phrases, but it diminishes towards their conclusion. A plausible explanation for this trend might be the heightened coordination requirements at the terminal points of phrases, especially as they transition to subsequent phrases and introduce fresh musical content [29]. This coordination challenge presents a contrasting scenario between polyphonic and homophonic textures. In the absence of a prominent leader, synchronisation in polyphonic settings tends to wane during these pivotal moments. Conversely, the presence of a melodic leader in homophonic compositions augments synchronisation, likely attributed to the augmented focus from the said leader. Future explorations employing eye-tracking methodologies might provide invaluable insights into these patterns, specifically by gauging visual interactions across different phrase segments [48; 138; 177].

Our first study exploited the *huSync* model to study dyadic synchronisation in a musical ensemble. Our research question addressed the influence of musical texture and phrase placement on interpersonal interactions, as discerned through visual cues tied to bodily movements. Interestingly, a subsequent analysis of the video's audio tracks unveiled a connection between coordination within the ensemble in terms of physical movements and auditory elements. Such a link aligns with previous research on ensemble synchronisation [133; 160; 178] and further underscores the expanding research supporting the multi-faceted aspects of musical communication [28; 154; 179]. This analysis also reinforces the significance of amalgamating visual and auditory indicators in understanding synchronisation within musical collectives. Our results indicate that *huSync* adeptly discerns dynamic alterations in interpersonal coupling, particularly those associated with ambiguities in leadership and coordination necessities, within standard video recordings of spontaneous small-group interactions.

In the second study, we extended *huSync* to test for directionality of influence and performed a pilot study utilising a subset of the same dataset as used in Study 1, making use of binary values [145; 146]. This study utilised Granger Causality, widely used in Neuroscience and Psychology, to measure information transfer and functional connectivity [180]. Our findings revealed that homophonic textures, characterised by a melodic leader and accompanying parts, exhibited more significant directional influence than polyphonic textures with a more balanced distribution of melodic content.

Transitioning to Study 3, while the previous exploration was a pilot based on a dataset subset, the third study delved deeper by encompassing the entire dataset to contextualise the insights from Study 2 further. Here, we deduced that homophonic textures consistently registered a higher frequency of significant GC values, suggesting a more substantial directional influence among pairs of co-performers. This consolidated our hypothesis that distinct roles in homophonic textures, akin to leader-follower dynamics, culminate in increased directionality in interpersonal coupling. Interestingly, in homophonic settings, the melody instrument consistently showcased a stronger influence over other instruments, reinforcing the idea of the melody player as the leader. These insights not only built upon our preliminary observations on non-directional coupling [13] but also resonated with existing studies that employed GC to scrutinise leadership dynamics in various ensemble configurations [5; 6; 7; 116; 181; 182]. In this investigation, we did not directly ascertain distinct leadership roles from the musicians but emphasised the intrinsic musical structure. A potential avenue for subsequent research is examining the extent to which homophonic textures allow for more liberal head movements due to diminished synchronisation requisites. While this study foregrounds head movements, it is imperative to acknowledge that they are not the exclusive, nor necessarily the paramount, metric for synchronisation in ensemble contexts. Our work underscored the role of musical texture in these dynamics and validated the capability of extracting leader-follower patterns from body motions in naturalistic video recordings. Our findings confirm hypothesis (H1) that computational and statistical techniques, such as phase-locking values and Granger Causality, are reliable measures of interpersonal synchronisation and leadership dynamics in small-group interactions.

7.1 Connection between Musical Texture and Leadership Dynamics

Pertinent to studies 2 and 3, the influence of musical texture on interpersonal coupling reveals that homophonic textures lead to more pronounced directionality, where information predominantly flows from the melody player to the accompanying players and less so the other way around. These findings confirm our results from Study 1 that the required leader-follower dynamics inherent in specific musical structures may allow for flexibility in interpersonal coupling. This adaptability can manifest in temporal adaptation and anticipation asymmetries during rhythmically coordinated interactions, a process that balances real-time error correction and future event prediction [4; 29]. Instrumental movement studies underscore that followers exhibit higher levels of adaptation and anticipation compared to leaders [135; 136; 183; 184]. Results from Studies 2 and 3 further suggest these asymmetries extend to additional body movements, implying that musical communication spans multiple timescales and sensory channels [47], consistent with diverse communication patterns observed elsewhere [185; 186].

However, it remains to be seen if other musical structural elements, such as phrase position, further modulate the interplay between musical texture and leadership dynamics. Results from Study 1 indicated a varying effect of texture based on phrase position. During initial and mid-phrase sections, coupling appears firmer for polyphonic versus homophonic textures but converges at phrase conclusions. The assertion is that polyphonic textures tend towards leader-follower dynamics at phrase endings, potentially due to coordination challenges [35]. Such complexities are evident in variable silent pauses that punctuate successive phrases, prompting increased communicative gestures like eye contact and pronounced movements [138; 177; 182]. Notably, leaders often adopt predictable timing strategies during these moments [42; 175; 187; 188], alluding to a broader principle of coordination smoothers [189]. Even though our current exploration did not delve into the influence of phrase position due to time series length constraints (refer to 7.3), future research may examine patterns of information flow among accompanying players. It could be especially revealing in this context, given our finding that such interpersonal dynamics are present and measurable.

Additionally, whether the hierarchical distinction in homophonic textures, distinguishing melody from accompaniment, inherently triggers leader-follower dynamics remains an open question. Past investigations involving piano duets demonstrated consistent melody leadership in keystroke timings and body sway [42; 160]. However, discerning whether this behaviour arises spontaneously or is a conscious strategy remains challenging. Interestingly, auditory leader-follower perceptions appear to be driven by automatic processing effects [133], whereas deliberate leadership actions are evident in head movements [116]. Our findings hint at implicit leadership dynamics that arise due to musical constraints. While pinpointing such dynamics remains elusive in controlled settings [190], future studies could potentially harness natural interactions, such as during rehearsal discussions, as a rich source of insights [191].

Finally, regardless of the emergence pattern of leader-follower dynamics, their impact on information directionality seems intertwined with attention dynamics. Ensemble players, committed to "prioritised integrative attending," distribute attention among their roles while encompassing the collective ensemble's sound [40]. Particularly in homophonic textures, while performers prioritise their part, there is a possible disproportionate attentional shift towards the melody by accompanying players [192]. This skewed attention distribution could drive directional interpersonal coupling dynamics, given the tight coupling between attention and sensory-motor processes [29; 193; 194].

7.2 Benefits of studying natural coordination

Our study adds to the growing literature showing that body movement, including head motion, provides an valid metric to investigate interpersonal coordination and leadership dynamics in group settings [6; 47; 116]. Moreover, we demonstrate the potential of markerless motion capture technology to analyse such leader-follower dynamics in videos of music ensemble performances recorded during live concerts. Studying such performances is informative as they offer a naturalistic setting for musical communication. While this context presents an ecologically valid representation of ensemble dynamics, it may not be the only approach to investigating leadership in musical performances. Research on music and dance underscores the influence of performing *in situ* on communication quality. Factors like acoustics and audience presence can impact performers' levels of motivation, attention, and arousal [195; 196; 197]. The nonverbal communication of emotions is a specific aspect that may benefit under these conditions, and the degree of enhancement may be reflected in increased information flow. Evidence for this link is seen in a study that found greater information flow in body sway when a trio were instructed to perform pieces with emotional expression than without emotion [3]. Heightened expressive intensity may, therefore, be associated with greater amplitude movements [198; 199] and the transmission of these cues between co-performers.

An additional advantage of using conventional video of ensemble performances is that it widens the potential pool of materials for analysis. Video-based analysis allows performances of other cultures to be studied when more specialised motion capture setups are not feasible [154], and therefore has the potential to overcome the prevailing WEIRD (Western, Educated, Industrialized, Rich, and Democratic societies) focus of research in psychology and neuroscience and subdisciplines such as music science [200]. A related practical benefit is that video is relatively neutral regarding group size, notwithstanding the issue of occlusion in large groups [13], and thus may help to accelerate the trend in the field to go beyond dyadic coordination to study groups of three or more performers [201]. An advantage of this upscaling, highlighted by the results in Study 3, is that it allowed us to examine the coupling between instrumentalists playing accompanying parts and interactions between melody and accompaniment. Our finding that the "other on other" influence was almost as strong as the "melody on other" influence in homophonic textures is noteworthy to the extent it captures the interaction between accompanying players. This finding suggests that to understand musical group dynamics, it is important to consider the interconnected network of the entire ensemble, in which subsets of performers function with some degree of independence [202]. Such independence is consistent with claims that it is necessary to balance the integration and segregation of psychological representations of "self" and "other" in ensemble performance [203; 204; 205], with the added nuance that there may be relatively high segregation between melody and accompaniment players, but a high degree of integration among the accompaniment players.

A further benefit of the current approach is conceptual. The observed effects of musical texture on leadership dynamics were not the result of an explicit experimental manipulation, nor were they post-hoc or data-driven findings. While these alternative approaches have particular strengths, experimental methods can lack ecological validity and data-driven findings can be challenging to interpret due to multiple possible contributing factors. Instead, our approach of segmenting the videoed performances based on musicological analysis of musical structure presents a middle ground that balances ecological naturalness and experimental control considerations. This feature highlights the benefit of using ensemble music from notated traditions, where the score functions as a script that constrains the actions of each performer while maintaining a degree of freedom for individual expression, as a domain to study the psychological dynamics of social interaction [12; 206].

7.3 Limitations

The *huSync* model, while generally reliable, faces various challenges. Below, we discuss two macro-level limitations: technical and methodological.

1. **Methodological Limitations**:

   (a) **Multiple Temporal Scales**: This study operates primarily at one temporal scale, focusing on head movements. However, synchronisa-

Figure 7.1: Performance images of the String Quartet No. 2 composed by Alexander Borodin (Top Left) and a trio for clarinet, viola, and piano composed by Robert Schumann (Bottom Left), with overlayed keypoints depicted on the right. These images illustrate the complexities and variations in pose estimation during different performances.

tion can occur at varied temporal scales, ranging from milliseconds to hours, leaving its exploration an open research question. Understanding such scales in cohesion and directionality is pivotal for revealing different aspects of synchronisation, a methodology unaddressed in the current study. Future research can investigate synchronisation across different temporal scales to provide more comprehensive insights.

(b) **Granger Causality (GC)**: GC is employed for statistical inquiries in this study, but its reliability for establishing causal relationships is still debated. The need to transform non-stationary data to stationary forms to apply GC might alter the intrinsic nature of the time series, potentially introducing biases [14; 207]. More robust methodologies are needed to establish causality without compromising data integrity.

(c) **Nuanced Roles of Head Movements in Ensemble Coordination**: Head movements, such as nods, have been linked to specific functions like timekeeping in ensemble music performance [138; 208].

Distinguishing these in future *huSync* iterations can provide richer performance analysis. Also, embedding the emotional encoding of head movements, currently absent, can yield a more comprehensive view of ensemble coordination [209; 210]. While this study foregrounds head movements, it is imperative to acknowledge that they are not the exclusive, nor necessarily the paramount, metric for synchronisation or directionality in ensemble contexts.

(d) **Generality of findings**: The study's focus on specific ensembles and the directionality between specific musical textures and performance dynamics might limit the general applicability of our findings. For instance, the interaction between melody and rhythm can significantly impact synchronisation, warranting a more detailed investigation. Exploring these findings in various non-musical scenarios, such as team sports or organisational settings where leadership dynamics are crucial, can offer valuable insights. Here, the study's implications can be leveraged to understand social signals and leadership dynamics in settings where one leads and others follow.

(e) **Control group**: Another limitation to address relates to the fact that, the present studies focused on naturalistic ensemble performance, and did not compare co-performers' body motion to movement patterns during solo performance. Previous research has highlighted the relevance of body movements in solo performance for regulating performance and communicating musical structure and expression [198; 211]. Incorporating a solo performance condition in future studies could provide a baseline for understanding how these functions are fulfilled for different musical textures in ensemble settings [212; 213]. It also bears mention that did not directly ascertain distinct leadership roles from the musicians but inferred them from the intrinsic musical structure. This means that they could be influenced by technical demands related to coordination, which varies as a function of the actual music being played. A potential avenue for subsequent research is examining the extent to which homophonic textures allow for more liberal head movements due to diminished synchronisation demands.

Note: Future works will consider exploring different temporal scales and robust causality methodologies, including applying findings in varied contexts to enhance the generality and applicability of the results.

2. **Technical Limitations**:

   (a) **Data Quality and Pose Estimation**: While the *huSync* technique offers several advantages, it faces challenges compared to traditional marker-based systems, which, albeit more expensive and potentially uncomfortable, yield less noisy data. Clear differentiation between video foreground and background is crucial for optimising pose estimation, a condition hard to meet in crowded frames with multiple individuals and obstructions, complicating analysis (See Fig. 7.1) [214]. Potential inaccuracies in pose estimation can distort synchronisation analysis outcomes, necessitating improvements in data acquisition and processing techniques.

   (b) **Technical Sensitivities**: The model is sensitive to various factors, including:

      i. The *huSync* system, as currently designed, does not differentiate between lateral (side-to-side) and vertical (up-down) head movements. It leverages input from a multi-person HPE algorithm, and, based on the musicians' arrangement within the scene, it solely utilises position time series along the x-axis (horizontal). However, with advancements in the HPE domain enabling data extraction in 3D space, there is potential for enhanced analysis by capturing more nuanced motion trajectories of participants in multiple directions.

ii. Challenges exist in tracking in larger ensembles or multi-row settings due to occlusions. Multi-camera setups and 3D viewpoints can be potential solutions [215; 216; 217], and while they show promise, refinement is inevitable for in-the-wild implementations.

iii. Lighting variations, frame resolution, and the lack of robust, open-source models trained on extensive datasets affecting output reliability.

(c) **Limited Test Cases**: The study is based on two specific instances of ensembles, limiting the scope of our understanding. Analysing a more diverse range of ensembles is essential for broader insights. Future research should consider different ensemble types and performance contexts to validate and extend the findings of this study.

3. **Others**: Studies indicate that performer familiarity significantly influences ensemble coordination and goal alignment. Knowledge of a co-performer part aids in coordinating head motion and body sway, though not necessarily the actual playing, serving as a mechanism to reconcile incongruent performance goals through stylistic assimilation [190]. Over time, as performers become more acquainted with each other, the overall flow of information may decrease, and relations may alter, with ensemble members aligning on a common stylistic interpretation and synchronising their body movements. This suggests a transition from dependency on feedback to adopting feedforward processes driven by internal models, impacting not

only the cohesion within the ensemble but also the interpretative dynamics of the performance [205; 218; 219; 220].

Chapter 8

Conclusions

This PhD thesis illustrates a comprehensive exploration into the computational modelling of synchronisation and leadership dynamics in small group interactions. Through the studies presented, we have contributed to bridge the gap between the intricate nuances of human interactions and the precision of computational modelling. This thesis addresses a notable challenge in studying interpersonal synchronisation and leadership dynamics within small group interactions. Traditionally, these domains' research has relied heavily on MoCap systems. While precise, these systems are expensive and intrusive, potentially altering the study participants' natural behaviour. We propose an alternative approach to examine small-group interactions in a more naturalistic setting: utilising non-intrusive human pose estimation algorithms. This approach led to the development of *huSync*, a model and framework that provides researchers with a more accessible and non-intrusive method of studying these interactions. Music ensembles are imbued with nonverbal communication, whether explicitly or implicitly, as participants naturally communicate through their body movements and gestures. Given this, we utilise music ensembles as a test-bed for conducting our experiments and validating the *huSync* model.

We explored various research questions when we established this model. One of the fundamental inquiries centred around musical phrases comparable to speech sentences. These phrases serve as coherent units within a musical composition. Based on existing literature, we hypothesised that musical texture and phrase position would influence the strength of interpersonal coupling in body motion. However, the impact of these musical elements on coupling strength at different phrase positions remained an open question. Another research question that emerged was to investigate how the effects of musical texture can affect leadership dynamics and the directionality of influence.

To empirically interrogate these questions, we operationalised the *huSync* model to methodically analyse chamber music concert recordings. This rigorous analysis proffered insights into the nexus between audio performance metrics and coupling measures, thereby enhancing our epistemological understanding of interpersonal synchronisation and the directionality of influence within musical ensembles. Within the purview of the extant literature, and to the best of our knowledge, our research represents first steps and being amongst the earliest to exploit HPE algorithms in extracting motoric and postural data from standardised video recordings, especially with an emphasis on dyadic synchronisation, leadership dynamics, and directionality of influence within a musical ensemble setup.

In Study 1 [13], we introduced *huSync*, designed to quantify and evaluate interpersonal synchronisation in small-group dyads. This study was pivotal in establishing a foundation for the subsequent investigations. The *huSync* model, with its multi-modal signal processing and feature extraction capabilities, provided a versatile and robust platform to analyse the subtleties of human interactions, particularly in musical ensembles.

With Study 2 [14], we delved deeper into the realm of musical interactions, focusing on the directionality of influence among ensemble musicians. By employing Granger causality, this study illuminated the intricate leader-follower dynamics that emerge during musical performances. The insights from this study underscored the importance of understanding the non-verbal cues and ancillary body movements that musicians employ to communicate and synchronise with one another.

In Study 3 [15], we further expanded on the theme of leadership dynamics, emphasising the quantifiable effects of musical structure on the directionality of influence in concert performance videos. The study explored how musical structure, particularly the relative salience of ensemble parts, can influence leadership dynamics. The study's results have provided strong evidence that how musicians interact with each other through their body movements mirrors the hierarchical relationship between the different parts of the musical ensemble. Such findings suggest that the directionality of influence between co-performers plays a crucial role in how the process of music creation evolves.

Lastly, Study 4 [16] extended beyond the core objectives, examining the correlation between dance and electroacoustic music phrases. This pilot study broadened our research scope, hinting at our methodology's potential applications in other forms of artistic expression.

8.1 Future works

The research lays the foundation for advances in interaction design and human dynamics through the *huSync* model. The following are projected future directions, and these can be worked upon in the framework of the upcoming PNRR project:

1. **Embodied Interaction Design**: An exploration into embodied interaction design, utilising the *huSync* model, can yield profound insights into the intricacies of non-verbal communication and interaction. This approach

seeks to design interactive systems more harmonious with human somatic states, paving the way for creating more intuitive, responsive interfaces and interaction paradigms.

2. **Multi-Temporal Scale Investigation**: The present study is constricted to a single temporal scale, predominantly examining head movements. However, synchronisation is a multifaceted phenomenon, unfolding across a spectrum of temporal scales—from milliseconds to hours. This vast and varied temporal landscape remains uncharted mainly in the current study, opening avenues for expansive research inquiries. Investigating how synchronisation manifests, interweaves, and directs across these different scales is critical for unveiling its diverse facets and implications. For instance, understanding how the immediate, millisecond-level synchrony of musicians' movements coexists and interacts with the overarching, hour-long harmonisation within a performance can offer a more holistic perspective on the intricate tapestry of synchronisation dynamics. Future research can uncover insights into temporal scales, synchronisation, filling knowledge gaps and extending understanding.

3. **Methodological and Technical Improvements**: Researchers are working on increasing the availability of more annotated data which will enhance the accuracy of HPE algorithms, allowing for the acquisition of more reliable motor and postural data under varied scene conditions, such as in low-light and crowded frames. Such improvements will extend the model's applicability and relevance in varied contexts in naturalistic settings.

4. **Cross-modal synchronisation analysis**: We are exploring methods to extend *huSync*'s applicability to cross-modal synchronisation analysis, such as measuring synchronisation between biophysical indicators, motor and

postural data from video and other data such as localisation of subject. This is an open area for investigation under PNRR projects such as proHome under Spoke 2.

5. **Broader research contexts and applications**: Extending the methodologies and findings to broader contexts can foster advancements in fields like human-computer interaction and social dynamics analysis. For example, analysing leader-follower dynamics in other group settings like orchestras or dance ensembles could broaden our understanding of social and leadership dynamics while continuing to use music as a test bed.

In synthesising the findings from these studies, a few overarching themes emerge. Firstly, the relationship of leadership and synchronisation in small group interactions, whether musical or otherwise, is an interplay of non-verbal cues, shared intentions, and mutual understanding. Secondly, the proposed computational model *huSync* offers a versatile and robust approach to studying these interactions, providing methodological rigour and empirical insights.

During this research, we systematically investigate the nuances of human interactions through a computational lens. Using video-based pose estimation to analyse large publicly available datasets of real-world performances across cultures and group sizes will maximise the potential use of music to study the communication dynamics of social groups. The methodologies and insights presented offer future researchers a scaffold to build further, helping enrich our shared comprehension of leader-follower relationships in small-group interactions.

Appendix A

Supplementary Analyses

A.1 Table for Brahms Quintet with bar numbers

Table A.1 indicates the bar numbers from the score corresponding to the Brahms Quintet. The “Concert Part” column indicates the specific part of Brahms’s Concert being analysed, such as “Part1”, “BrahmsConcertPart1” and so on. The “No.” column represents a sequence number given to each analysed segment within the specified part. The “Start Bar” and “End Bar” columns show the range of musical bars included in each analysed segment, essentially telling us where in the composition the segment begins and ends. The columns “Start Time” and “End Time” indicate the timings in seconds, at which the respective musical bars begin and end. The ”Duration Difference” column calculates the time span of each segment by subtracting the “Start Time” from the ”End Time,” giving a clear picture of how long each segment lasts. The “No.Instr.” column indicates how many instruments are being played in each segment, contributing to the understanding of the complexity of the musical composition. The “T” column provides insights into the texture of the musical piece. In this case, ‘P’ stands for ‘Polyphonic’ and ‘H’ stands for ‘Homophonic’. The “MI” column (or Melody Instrument) identifies the instrument playing the main melody in each segment.

The abbreviations correspond to specific instruments, for instance, 'Cl' represents the clarinet, and 'Vln1' represents the first violin. In segments where the melody is shared or alternates between two instruments, both are listed, separated by a slash.

A.2 Table for Borodin Quartet with bar numbers

Table A.2 indicates the bar numbers from the score corresponding to the Borodin Quartet. The "Concert Part" column indicates the specific part of Borodin's work being analysed, such as "Part 1", "Part 2" and so on. The "No." column represents a sequence number given to each analysed segment within the specified part. The "Start Bar" and "End Bar" columns show the range of musical bars included in each analysed segment, essentially telling us where in the composition the segment begins and ends. The columns "Start Time" and "End Time" indicate the timings in seconds, at which the respective musical bars begin and end. The "Duration Difference" column calculates the time span of each segment by subtracting the "Start Time" from the "End Time", giving a clear picture of how long each segment lasts. The "No.Instr." column indicates how many instruments are being played in each segment, contributing to the understanding of the complexity of the musical composition. The "T" column provides insights into the texture of the musical piece. In this case, 'P' stands for 'Polyphonic' and 'H' stands for 'Homophonic'. The "MI" column (or Melody Instrument) identifies the instrument playing the main melody in each segment. 'NA' indicates that there isn't a single main melody instrument but rather is rather distributed. Other abbreviations correspond to specific instruments, such as 'Vln1' for first violin, 'Vlc' for cello, 'Vla' for viola, and 'Vln2' for second violin.

Concert Part	No.	Start Bar	End Bar	Start Time	End Time	Duration Diff	No.Instr.	T	MI
Part1	1	5	12	12.66	31.33	18.670	5	P	Cl
Part1	3	37	47	96.97	124.67	27.702	5	H	Cl
Part1	4	51	57	132.36	153.07	20.710	5	H	Cl
Part1	5	5	12	184.03	201.87	17.837	5	P	Cl
Part1	7	37	47	267.25	295.24	27.991	5	H	Cl
Part1	8	51	57	304.25	324.77	20.519	5	H	Cl
Part1	10	87	97	397.70	425.01	27.312	5	P	Cl/Vln1
Part1	11	98	105	425.01	448.63	23.624	5	H	Cl/Vln1
Part1	13	114	120	472.38	491.91	19.537	5	P	Cl
Part1	16	162	168	599.76	616.91	17.151	5	P	Cl/Vln1
Part1	17	172	178	625.77	646.55	20.785	5	P	Vln1
Part1	19	199	205	694.02	710.19	16.161	5	H	Vln1
Part1	20	211	218	724.92	755.35	30.433	5	H	Cl
Part2	3	27	31	84.50	103.72	19.213	5	H	Cl
Part2	7	54	56	191.64	217.04	25.392	5	P	Cl
Part2	8	57	59	217.04	235.76	18.720	5	H	Cl
Part2	9	60	62	235.76	259.39	23.633	5	P	Cl
Part2	11	68	71	288.78	313.03	24.248	5	P	Cl
Part2	13	79	87	350.75	388.95	38.199	5	H	Cl/Vln1
Part2	14	88	93	388.95	416.50	27.553	5	P	Cl
Part2	17	113	117	489.83	508.93	19.102	5	H	Cl
Part3	3	121	128	83.78	99.27	15.488	5	P	Vln1/Cl
Part3	5	153	160	171.16	186.82	15.656	5	P	Vln1/Cl
Part3	7	218	222	310.91	333.40	22.488	5	H	Cl

Table A.1: Detailed data for Brahms concert parts.

Concert Part	No.	Start Bar	End Bar	Start Time	End Time	Duration Diff	No.Instr.	T	MI
Part1	1	9	21	13.099	34.37	21.271	4	P	NA
Part1	2	43	56	61.962	82.539	20.577	4	P	NA
Part2	1	380	395	262.474	277.616	15.142	4	P	NA
Part2	2	540	553	404.765	422.774	18.009	4	H	Vln1
Part2	3	554	569	422.774	441.398	18.624	4	H	Vln1
Part2	4	23	32	828.574	858.202	29.628	4	P	Vln1
Part2	5	33	40	858.202	881.197	22.995	4	P	Vlc
Part2	6	55	62	921.932	945.796	23.864	4	H	Vln1
Part2	7	76	82	984.981	1004.932	19.951	4	P	NA
Part2	8	97	103	1045.337	1066.676	21.339	4	P	NA
Part2	9	104	108	1066.676	1081.865	15.189	4	P	NA
Part3	1	118	125	19.706	45.783	26.077	4	P	NA
Part3	2	126	131	45.783	61.591	15.808	4	H	Vln1
Part4	1	163	168	28.183	46.895	18.712	4	H	Vln1
Part4	2	169	176	46.895	71.868	24.973	4	H	Vln1
Part4	3	316	324	258.159	273.454	15.295	4	H	Vln1
Part4	4	325	334	273.454	290.12	16.666	4	H	Vla
Part4	5	351	360	316.721	333.126	16.405	4	H	Vln2
Part4	6	67	79	637.503	660.318	22.815	4	H	Vln1
Part4	7	155	164	782.907	798.735	15.828	4	P	Vln1

Table A.2: Detailed data for Borodin concert parts.

A.3 Table for Dyadic Synchronisation test results in Brahms Quintet

The table below presents the results of the Dyadic Synchronisation tests carried out on different parts of the Brahms Quintet Quintet. The "Concert & FileNo." column identifies the specific concert and consequent section of the Quartet being analysed. The "T" column signifies the texture of the musical piece, with 'P' representing 'Polyphonic' and 'H' denoting 'Homophonic'. The "Pair" column shows the dyadic pairs of musicians being examined for their synchronisation, such as "P1_P2", "P2_P3", etc. The columns "Start", "Middle", and "End" are the specific dyadic synchronisation values obtained for the start, middle, and end of the musical phrases. They signify the phase-locking values of the specific musician pair. The "Duration" column is the duration of the entire musical section or phrase.

Table A.3: Detailed PLV data for Brahms Concert Parts.

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
Brahms Part 1_1	P	P1_P2	0.23	0.34	0.34	18.67
Brahms Part 1_1	P	P1_P3	0.21	0.34	0.28	18.67
Brahms Part 1_1	P	P1_P4	0.29	0.38	0.35	18.67
Brahms Part 1_1	P	P1_P5	0.32	0.28	0.31	18.67
Brahms Part 1_1	P	P2_P3	0.32	0.34	0.28	18.67
Brahms Part 1_1	P	P2_P4	0.35	0.38	0.33	18.67
Brahms Part 1_1	P	P2_P5	0.33	0.48	0.34	18.67
Brahms Part 1_1	P	P3_P4	0.37	0.33	0.27	18.67
Brahms Part 1_1	P	P3_P5	0.36	0.34	0.34	18.67
Brahms Part 1_1	P	P4_P5	0.40	0.34	0.29	18.67
Brahms Part 1_3	H	P1_P2	0.23	0.25	0.21	27.70
Brahms Part 1_3	H	P1_P3	0.20	0.25	0.23	27.70
Brahms Part 1_3	H	P1_P4	0.23	0.31	0.23	27.70

Continued on next page

Table A.3 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
Brahms Part 1_3	H	P1_P5	0.18	0.25	0.20	27.70
Brahms Part 1_3	H	P2_P3	0.21	0.14	0.19	27.70
Brahms Part 1_3	H	P2_P4	0.23	0.25	0.30	27.70
Brahms Part 1_3	H	P2_P5	0.14	0.18	0.22	27.70
Brahms Part 1_3	H	P3_P4	0.23	0.24	0.19	27.70
Brahms Part 1_3	H	P3_P5	0.20	0.23	0.26	27.70
Brahms Part 1_3	H	P4_P5	0.26	0.22	0.19	27.70
Brahms Part 1_4	H	P1_P2	0.18	0.18	0.19	20.71
Brahms Part 1_4	H	P1_P3	0.26	0.30	0.28	20.71
Brahms Part 1_4	H	P1_P4	0.36	0.28	0.37	20.71
Brahms Part 1_4	H	P1_P5	0.34	0.31	0.35	20.71
Brahms Part 1_4	H	P2_P3	0.29	0.26	0.31	20.71
Brahms Part 1_4	H	P2_P4	0.27	0.24	0.26	20.71
Brahms Part 1_4	H	P2_P5	0.22	0.24	0.23	20.71
Brahms Part 1_4	H	P3_P4	0.27	0.31	0.26	20.71
Brahms Part 1_4	H	P3_P5	0.30	0.28	0.29	20.71
Brahms Part 1_4	H	P4_P5	0.34	0.26	0.33	20.71
Brahms Part 1_5	P	P1_P2	0.30	0.35	0.28	17.84
Brahms Part 1_5	P	P1_P3	0.25	0.24	0.29	17.84
Brahms Part 1_5	P	P1_P4	0.34	0.36	0.33	17.84
Brahms Part 1_5	P	P1_P5	0.31	0.37	0.27	17.84
Brahms Part 1_5	P	P2_P3	0.35	0.30	0.26	17.84
Brahms Part 1_5	P	P2_P4	0.31	0.41	0.46	17.84
Brahms Part 1_5	P	P2_P5	0.41	0.27	0.39	17.84
Brahms Part 1_5	P	P3_P4	0.32	0.26	0.32	17.84
Brahms Part 1_5	P	P3_P5	0.41	0.34	0.39	17.84
Brahms Part 1_5	P	P4_P5	0.29	0.33	0.29	17.84
Brahms Part 1_7	H	P1_P2	0.23	0.19	0.28	27.99
Brahms Part 1_7	H	P1_P3	0.18	0.21	0.30	27.99
Brahms Part 1_7	H	P1_P4	0.18	0.23	0.26	27.99
Brahms Part 1_7	H	P1_P5	0.21	0.23	0.33	27.99

Continued on next page

Table A.3 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
Brahms Part 1_7	H	P2_P3	0.25	0.22	0.28	27.99
Brahms Part 1_7	H	P2_P4	0.21	0.24	0.26	27.99
Brahms Part 1_7	H	P2_P5	0.19	0.21	0.26	27.99
Brahms Part 1_7	H	P3_P4	0.20	0.25	0.26	27.99
Brahms Part 1_7	H	P3_P5	0.27	0.22	0.28	27.99
Brahms Part 1_7	H	P4_P5	0.26	0.26	0.30	27.99
Brahms Part 1_8	H	P1_P2	0.29	0.32	0.26	20.52
Brahms Part 1_8	H	P1_P3	0.24	0.34	0.32	20.52
Brahms Part 1_8	H	P1_P4	0.27	0.29	0.34	20.52
Brahms Part 1_8	H	P1_P5	0.22	0.34	0.30	20.52
Brahms Part 1_8	H	P2_P3	0.16	0.38	0.30	20.52
Brahms Part 1_8	H	P2_P4	0.29	0.31	0.25	20.52
Brahms Part 1_8	H	P2_P5	0.28	0.29	0.28	20.52
Brahms Part 1_8	H	P3_P4	0.30	0.26	0.33	20.52
Brahms Part 1_8	H	P3_P5	0.27	0.19	0.30	20.52
Brahms Part 1_8	H	P4_P5	0.18	0.26	0.34	20.52
Brahms Part 1_10	P	P1_P2	0.32	0.41	0.27	27.31
Brahms Part 1_10	P	P1_P3	0.34	0.25	0.26	27.31
Brahms Part 1_10	P	P1_P4	0.33	0.22	0.30	27.31
Brahms Part 1_10	P	P1_P5	0.44	0.31	0.21	27.31
Brahms Part 1_10	P	P2_P3	0.25	0.38	0.26	27.31
Brahms Part 1_10	P	P2_P4	0.20	0.36	0.22	27.31
Brahms Part 1_10	P	P2_P5	0.25	0.61	0.32	27.31
Brahms Part 1_10	P	P3_P4	0.19	0.27	0.25	27.31
Brahms Part 1_10	P	P3_P5	0.37	0.39	0.28	27.31
Brahms Part 1_10	P	P4_P5	0.24	0.34	0.24	27.31
Brahms Part 1_11	H	P1_P2	0.54	0.42	0.46	23.62
Brahms Part 1_11	H	P1_P3	0.40	0.25	0.22	23.62
Brahms Part 1_11	H	P1_P4	0.34	0.18	0.21	23.62
Brahms Part 1_11	H	P1_P5	0.15	0.28	0.22	23.62
Brahms Part 1_11	H	P2_P3	0.44	0.42	0.21	23.62

Continued on next page

Table A.3 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
Brahms Part 1_11	H	P2_P4	0.39	0.26	0.25	23.62
Brahms Part 1_11	H	P2_P5	0.26	0.21	0.18	23.62
Brahms Part 1_11	H	P3_P4	0.64	0.48	0.48	23.62
Brahms Part 1_11	H	P3_P5	0.26	0.15	0.17	23.62
Brahms Part 1_11	H	P4_P5	0.40	0.23	0.57	23.62
Brahms Part 1_13	P	P1_P2	0.36	0.50	0.33	19.54
Brahms Part 1_13	P	P1_P3	0.35	0.48	0.26	19.54
Brahms Part 1_13	P	P1_P4	0.35	0.42	0.27	19.54
Brahms Part 1_13	P	P1_P5	0.29	0.46	0.26	19.54
Brahms Part 1_13	P	P2_P3	0.29	0.57	0.24	19.54
Brahms Part 1_13	P	P2_P4	0.27	0.47	0.23	19.54
Brahms Part 1_13	P	P2_P5	0.19	0.49	0.23	19.54
Brahms Part 1_13	P	P3_P4	0.21	0.36	0.22	19.54
Brahms Part 1_13	P	P3_P5	0.38	0.47	0.34	19.54
Brahms Part 1_13	P	P4_P5	0.25	0.29	0.32	19.54
Brahms Part 1_16	P	P1_P2	0.26	0.31	0.30	17.15
Brahms Part 1_16	P	P1_P3	0.27	0.26	0.27	17.15
Brahms Part 1_16	P	P1_P4	0.23	0.28	0.20	17.15
Brahms Part 1_16	P	P1_P5	0.29	0.24	0.34	17.15
Brahms Part 1_16	P	P2_P3	0.31	0.27	0.42	17.15
Brahms Part 1_16	P	P2_P4	0.18	0.24	0.31	17.15
Brahms Part 1_16	P	P2_P5	0.39	0.33	0.23	17.15
Brahms Part 1_16	P	P3_P4	0.30	0.27	0.31	17.15
Brahms Part 1_16	P	P3_P5	0.33	0.30	0.26	17.15
Brahms Part 1_16	P	P4_P5	0.35	0.30	0.25	17.15
Brahms Part 1_17	P	P1_P2	0.23	0.26	0.28	20.79
Brahms Part 1_17	P	P1_P3	0.30	0.26	0.35	20.79
Brahms Part 1_17	P	P1_P4	0.23	0.23	0.21	20.79
Brahms Part 1_17	P	P1_P5	0.30	0.32	0.26	20.79
Brahms Part 1_17	P	P2_P3	0.37	0.28	0.29	20.79
Brahms Part 1_17	P	P2_P4	0.34	0.22	0.28	20.79

Continued on next page

Table A.3 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
Brahms Part 1_17	P	P2_P5	0.33	0.38	0.29	20.79
Brahms Part 1_17	P	P3_P4	0.37	0.24	0.34	20.79
Brahms Part 1_17	P	P3_P5	0.33	0.24	0.37	20.79
Brahms Part 1_17	P	P4_P5	0.37	0.33	0.28	20.79
Brahms Part 1_18	H	P1_P2	0.22	0.30	0.25	17.11
Brahms Part 1_18	H	P1_P3	0.25	0.27	0.25	17.11
Brahms Part 1_18	H	P1_P4	0.27	0.31	0.31	17.11
Brahms Part 1_18	H	P1_P5	0.32	0.34	0.30	17.11
Brahms Part 1_18	H	P2_P3	0.26	0.33	0.30	17.11
Brahms Part 1_18	H	P2_P4	0.23	0.23	0.27	17.11
Brahms Part 1_18	H	P2_P5	0.32	0.32	0.25	17.11
Brahms Part 1_18	H	P3_P4	0.35	0.19	0.21	17.11
Brahms Part 1_18	H	P3_P5	0.32	0.26	0.26	17.11
Brahms Part 1_18	H	P4_P5	0.24	0.29	0.30	17.11

A.4 Table for Dyadic Synchronisation test results in Borodin Quartet

The table below presents the results of the Dyadic Synchronisation tests carried out on different parts of the Borodin Quartet Quintet. The “Concert & FileNo.” column identifies the specific concert and consequent section of the Quartet being analysed. The “T” column signifies the texture of the musical piece, with ’P’ representing ’Polyphonic’ and ’H’ denoting ’Homophonic’. The “Pair” column shows the dyadic pairs of musicians being examined for their synchronisation, such as “P1_P2”, “P2_P3”, etc. The columns “Start”, “Middle”, and “End” are the specific dyadic synchronisation values obtained for the start, middle, and end of the musical phrases. They signify the phase-locking values of the specific musician pair. The “Duration” column is the duration of the entire musical section or phrase.

A.4 Table for Dyadic Synchronisation test results in Borodin Quartet

Table A.4: Detailed PLV data for Borodin Concert Parts.

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
BorodinPart1_1	P	P1_P2	0.39	0.28	0.31	21.27
BorodinPart1_1	P	P1_P3	0.33	0.33	0.21	21.27
BorodinPart1_1	P	P1_P4	0.28	0.20	0.21	21.27
BorodinPart1_1	P	P2_P3	0.30	0.27	0.23	21.27
BorodinPart1_1	P	P2_P4	0.25	0.26	0.27	21.27
BorodinPart1_1	P	P3_P4	0.25	0.24	0.18	21.27
BorodinPart1_2	P	P1_P2	0.27	0.28	0.32	20.58
BorodinPart1_2	P	P1_P3	0.28	0.30	0.29	20.58
BorodinPart1_2	P	P1_P4	0.36	0.31	0.31	20.58
BorodinPart1_2	P	P2_P3	0.32	0.24	0.29	20.58
BorodinPart1_2	P	P2_P4	0.33	0.27	0.32	20.58
BorodinPart1_2	P	P3_P4	0.28	0.27	0.31	20.58
BorodinPart2_1	P	P1_P2	0.23	0.43	0.25	15.14
BorodinPart2_1	P	P1_P3	0.36	0.39	0.29	15.14
BorodinPart2_1	P	P1_P4	0.32	0.32	0.35	15.14
BorodinPart2_1	P	P2_P3	0.28	0.32	0.26	15.14
BorodinPart2_1	P	P2_P4	0.23	0.38	0.38	15.14
BorodinPart2_1	P	P3_P4	0.27	0.31	0.27	15.14
BorodinPart2_2	H	P1_P2	0.32	0.28	0.30	18.01
BorodinPart2_2	H	P1_P3	0.28	0.30	0.33	18.01
BorodinPart2_2	H	P1_P4	0.30	0.34	0.23	18.01
BorodinPart2_2	H	P2_P3	0.28	0.31	0.28	18.01
BorodinPart2_2	H	P2_P4	0.27	0.22	0.36	18.01
BorodinPart2_2	H	P3_P4	0.29	0.25	0.31	18.01
BorodinPart2_3	H	P1_P2	0.28	0.41	0.27	18.62
BorodinPart2_3	H	P1_P3	0.19	0.31	0.29	18.62
BorodinPart2_3	H	P1_P4	0.26	0.29	0.27	18.62
BorodinPart2_3	H	P2_P3	0.27	0.28	0.36	18.62
BorodinPart2_3	H	P2_P4	0.24	0.24	0.26	18.62
BorodinPart2_3	H	P3_P4	0.26	0.30	0.24	18.62

Continued on next page

Table A.4 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
BorodinPart2_4	P	P1_P2	0.22	0.19	0.18	29.63
BorodinPart2_4	P	P1_P3	0.27	0.27	0.24	29.63
BorodinPart2_4	P	P1_P4	0.23	0.20	0.17	29.63
BorodinPart2_4	P	P2_P3	0.27	0.25	0.17	29.63
BorodinPart2_4	P	P2_P4	0.24	0.22	0.25	29.63
BorodinPart2_4	P	P3_P4	0.28	0.17	0.21	29.63
BorodinPart2_5	P	P1_P2	0.23	0.23	0.37	23.00
BorodinPart2_5	P	P1_P3	0.21	0.34	0.28	23.00
BorodinPart2_5	P	P1_P4	0.22	0.26	0.24	23.00
BorodinPart2_5	P	P2_P3	0.30	0.26	0.28	23.00
BorodinPart2_5	P	P2_P4	0.29	0.21	0.33	23.00
BorodinPart2_5	P	P3_P4	0.29	0.27	0.28	23.00
BorodinPart2_6	H	P1_P2	0.26	0.33	0.23	23.86
BorodinPart2_6	H	P1_P3	0.31	0.27	0.29	23.86
BorodinPart2_6	H	P1_P4	0.16	0.20	0.25	23.86
BorodinPart2_6	H	P2_P3	0.28	0.21	0.33	23.86
BorodinPart2_6	H	P2_P4	0.28	0.26	0.26	23.86
BorodinPart2_6	H	P3_P4	0.35	0.22	0.30	23.86
BorodinPart2_7	P	P1_P2	0.37	0.39	0.33	19.95
BorodinPart2_7	P	P1_P3	0.26	0.34	0.28	19.95
BorodinPart2_7	P	P1_P4	0.35	0.25	0.29	19.95
BorodinPart2_7	P	P2_P3	0.33	0.36	0.31	19.95
BorodinPart2_7	P	P2_P4	0.33	0.26	0.35	19.95
BorodinPart2_7	P	P3_P4	0.30	0.33	0.29	19.95
BorodinPart2_8	P	P1_P2	0.19	0.27	0.22	21.34
BorodinPart2_8	P	P1_P3	0.28	0.25	0.31	21.34
BorodinPart2_8	P	P1_P4	0.34	0.40	0.32	21.34
BorodinPart2_8	P	P2_P3	0.24	0.34	0.30	21.34
BorodinPart2_8	P	P2_P4	0.30	0.36	0.29	21.34
BorodinPart2_8	P	P3_P4	0.26	0.36	0.24	21.34
BorodinPart2_9	P	P1_P2	0.45	0.31	0.33	15.19

Continued on next page

Table A.4 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
BorodinPart2_9	P	P1_P3	0.41	0.39	0.25	15.19
BorodinPart2_9	P	P1_P4	0.31	0.50	0.32	15.19
BorodinPart2_9	P	P2_P3	0.43	0.25	0.28	15.19
BorodinPart2_9	P	P2_P4	0.34	0.33	0.30	15.19
BorodinPart2_9	P	P3_P4	0.29	0.30	0.28	15.19
BorodinPart3_1	P	P1_P2	0.28	0.27	0.17	26.08
BorodinPart3_1	P	P1_P3	0.27	0.26	0.26	26.08
BorodinPart3_1	P	P1_P4	0.31	0.25	0.31	26.08
BorodinPart3_1	P	P2_P3	0.20	0.28	0.31	26.08
BorodinPart3_1	P	P2_P4	0.25	0.26	0.28	26.08
BorodinPart3_1	P	P3_P4	0.23	0.23	0.25	26.08
BorodinPart3_2	H	P1_P2	0.30	0.30	0.29	15.81
BorodinPart3_2	H	P1_P3	0.24	0.23	0.24	15.81
BorodinPart3_2	H	P1_P4	0.30	0.28	0.30	15.81
BorodinPart3_2	H	P2_P3	0.25	0.25	0.28	15.81
BorodinPart3_2	H	P2_P4	0.27	0.29	0.29	15.81
BorodinPart3_2	H	P3_P4	0.25	0.26	0.26	15.81
BorodinPart4_1	H	P1_P2	0.31	0.29	0.27	18.71
BorodinPart4_1	H	P1_P3	0.34	0.32	0.24	18.71
BorodinPart4_1	H	P1_P4	0.23	0.32	0.34	18.71
BorodinPart4_1	H	P2_P3	0.22	0.33	0.26	18.71
BorodinPart4_1	H	P2_P4	0.36	0.32	0.33	18.71
BorodinPart4_1	H	P3_P4	0.27	0.31	0.22	18.71
BorodinPart4_2	H	P1_P2	0.26	0.22	0.23	24.97
BorodinPart4_2	H	P1_P3	0.16	0.24	0.27	24.97
BorodinPart4_2	H	P1_P4	0.18	0.28	0.30	24.97
BorodinPart4_2	H	P2_P3	0.22	0.19	0.24	24.97
BorodinPart4_2	H	P2_P4	0.18	0.29	0.25	24.97
BorodinPart4_2	H	P3_P4	0.26	0.17	0.28	24.97
BorodinPart4_3	H	P1_P2	0.32	0.37	0.36	15.30
BorodinPart4_3	H	P1_P3	0.24	0.30	0.40	15.30

Continued on next page

Table A.4 – continued from previous page

Concert & FileNo	Texture	Pair	Start	Middle	End	Duration
BorodinPart4_3	H	P1_P4	0.30	0.48	0.39	15.30
BorodinPart4_3	H	P2_P3	0.26	0.42	0.36	15.30
BorodinPart4_3	H	P2_P4	0.28	0.33	0.37	15.30
BorodinPart4_3	H	P3_P4	0.26	0.32	0.34	15.30
BorodinPart4_4	H	P1_P2	0.24	0.27	0.26	16.67
BorodinPart4_4	H	P1_P3	0.30	0.22	0.27	16.67
BorodinPart4_4	H	P1_P4	0.29	0.38	0.41	16.67
BorodinPart4_4	H	P2_P3	0.32	0.29	0.31	16.67
BorodinPart4_4	H	P2_P4	0.35	0.30	0.26	16.67
BorodinPart4_4	H	P3_P4	0.29	0.31	0.21	16.67
BorodinPart4_5	H	P1_P2	0.28	0.29	0.23	16.41
BorodinPart4_5	H	P1_P3	0.25	0.25	0.33	16.41
BorodinPart4_5	H	P1_P4	0.29	0.29	0.29	16.41
BorodinPart4_5	H	P2_P3	0.27	0.28	0.39	16.41
BorodinPart4_5	H	P2_P4	0.29	0.29	0.32	16.41
BorodinPart4_5	H	P3_P4	0.26	0.26	0.29	16.41
BorodinPart4_6	H	P1_P2	0.28	0.29	0.25	22.82
BorodinPart4_6	H	P1_P3	0.22	0.20	0.25	22.82
BorodinPart4_6	H	P1_P4	0.29	0.26	0.25	22.82
BorodinPart4_6	H	P2_P3	0.26	0.28	0.18	22.82
BorodinPart4_6	H	P2_P4	0.28	0.23	0.28	22.82
BorodinPart4_6	H	P3_P4	0.24	0.26	0.22	22.82
BorodinPart4_7	P	P1_P2	0.42	0.24	0.47	15.83
BorodinPart4_7	P	P1_P3	0.29	0.29	0.24	15.83
BorodinPart4_7	P	P1_P4	0.32	0.34	0.30	15.83
BorodinPart4_7	P	P2_P3	0.25	0.34	0.23	15.83
BorodinPart4_7	P	P2_P4	0.25	0.40	0.33	15.83
BorodinPart4_7	P	P3_P4	0.37	0.34	0.35	15.83

A.5 Table for Granger Causality test results in Brahms Quintet

The table below presents the results of the Granger Causality tests carried out on different parts of the Brahms Quintet. The "Part" column identifies the specific section of the Quintet being analysed. The "File No." column provides a sequential numbering of the analysed segments within each part. The "M1" and "M2" columns represent the first and second musicians in a dyadic pair being analysed. The "T" column signifies the texture of the musical piece, with 'P' representing 'Polyphonic' and 'H' denoting 'Homophonic'. The "Pair" column shows the dyadic pairs of musicians being examined for their Granger causality, such as "m1_m2", "m1_m3", etc. The "F_M1_M2" and "F_M2_M1" columns provide the F values, which are statistical measures indicating the strength of causality from musician 1 to musician 2, and vice versa. The "p_M1_M2" and "p_M2_M1" columns present the corresponding p-values, which are probabilities used to determine the significance of the observed F values. The columns "p_M1_M2 (B)" and "p_M2_M1 (B)" hold binary values (1 or 0) indicating whether the pairs Granger causes each other or not, with 1 implying causality and 0 suggesting no causality. The "M_I" column signifies the main instrument involved in the Granger Causality analysis, providing additional details about the musicians' dyadic pair and the kind of instruments they played in each analysed segment. If a combination of instruments was involved, it is represented as "Mixed".

Table A.5: Detailed Granger Causality data for Brahms Concert Parts.

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part1	1	m1	m2	P	0.958	0.532	1.379	0.09	0	0	Mixed
Part1	1	m1	m3	P	0.648	0.926	1.132	0.291	0	0	Mixed
Part1	1	m1	m4	P	0.467	0.993	0.748	0.832	0	0	Mixed

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part1	1	m1	m5	P	0.487	0.991	0.949	0.546	0	0	Mixed
Part1	1	m2	m3	P	0.958	0.533	1.265	0.161	0	0	Mixed
Part1	1	m2	m4	P	0.982	0.495	1.328	0.118	0	0	Mixed
Part1	1	m2	m5	P	1.082	0.354	1.681	0.015	0	1	Mixed
Part1	1	m3	m4	P	1.697	0.013	1.182	0.236	1	0	Mixed
Part1	1	m3	m5	P	1.162	0.257	1.162	0.258	0	0	Mixed
Part1	1	m4	m5	P	0.6	0.955	1.173	0.245	0	0	Mixed
Part1	3	m1	m2	H	0.96	0.529	1.583	0.026	0	1	m5
Part1	3	m1	m3	H	1.359	0.097	1.67	0.014	0	1	m5
Part1	3	m1	m4	H	1.158	0.258	1.624	0.02	0	1	m5
Part1	3	m1	m5	H	1.383	0.085	0.981	0.496	0	0	m5
Part1	3	m2	m3	H	1.644	0.017	1.252	0.168	1	0	m5
Part1	3	m2	m4	H	1.392	0.081	1.878	0.003	0	1	m5
Part1	3	m2	m5	H	1.719	0.01	1.572	0.027	1	1	m5
Part1	3	m3	m4	H	1.888	0.003	1.181	0.234	1	0	m5
Part1	3	m3	m5	H	1.755	0.008	1.287	0.141	1	0	m5
Part1	3	m4	m5	H	1.436	0.063	1.085	0.347	0	0	m5
Part2	3	m1	m2	H	1.385	0.087	2.691	0	0	1	m5
Part2	3	m1	m3	H	1.847	0.005	2.656	0	1	1	m5
Part2	3	m1	m4	H	1.282	0.148	2.538	0	0	1	m5
Part2	3	m1	m5	H	1.056	0.389	2.567	0	0	1	m5
Part2	3	m2	m3	H	1.721	0.011	1.411	0.076	1	0	m5
Part2	3	m2	m4	H	1.767	0.008	1.568	0.03	1	1	m5
Part2	3	m2	m5	H	2.704	0	2.393	0	1	1	m5
Part2	3	m3	m4	H	1.626	0.021	1.628	0.021	1	1	m5
Part2	3	m3	m5	H	1.731	0.01	2.703	0	1	1	m5
Part2	3	m4	m5	H	2.415	0	2.117	0.001	1	1	m5
Part3	3	m1	m2	P	0.954	0.539	1.139	0.286	0	0	Mixed
Part3	3	m1	m3	P	0.827	0.729	2.173	0	0	1	Mixed
Part3	3	m1	m4	P	1.581	0.029	0.785	0.786	1	0	Mixed
Part3	3	m1	m5	P	0.805	0.759	1.455	0.061	0	0	Mixed

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part3	3	m2	m3	P	0.882	0.649	0.883	0.647	0	0	Mixed
Part3	3	m2	m4	P	1.653	0.019	1.191	0.23	1	0	Mixed
Part3	3	m2	m5	P	1.373	0.096	1.967	0.002	0	1	Mixed
Part3	3	m3	m4	P	1.465	0.058	0.811	0.751	0	0	Mixed
Part3	3	m3	m5	P	0.861	0.68	0.817	0.743	0	0	Mixed
Part3	3	m4	m5	P	0.709	0.873	1.085	0.351	0	0	Mixed
Part1	4	m1	m2	H	0.892	0.634	2.371	0	0	1	m5
Part1	4	m1	m3	H	1.346	0.107	1.184	0.233	0	0	m5
Part1	4	m1	m4	H	1.457	0.058	1.293	0.14	0	0	m5
Part1	4	m1	m5	H	0.914	0.601	2.426	0	0	1	m5
Part1	4	m2	m3	H	1.617	0.022	1.258	0.166	1	0	m5
Part1	4	m2	m4	H	0.697	0.887	1.984	0.002	0	1	m5
Part1	4	m2	m5	H	0.925	0.583	2.261	0	0	1	m5
Part1	4	m3	m4	H	1.283	0.147	1.76	0.008	0	1	m5
Part1	4	m3	m5	H	0.495	0.99	0.734	0.849	0	0	m5
Part1	4	m4	m5	H	0.504	0.988	1.335	0.113	0	0	m5
Part1	5	m1	m2	P	3.013	0	1.444	0.063	1	0	Mixed
Part1	5	m1	m3	P	1.334	0.115	2.194	0	0	1	Mixed
Part1	5	m1	m4	P	1.249	0.175	1.02	0.439	0	0	Mixed
Part1	5	m1	m5	P	1.511	0.043	1.775	0.008	1	1	Mixed
Part1	5	m2	m3	P	0.661	0.916	2.907	0	0	1	Mixed
Part1	5	m2	m4	P	1.587	0.027	1.036	0.416	1	0	Mixed
Part1	5	m2	m5	P	3.682	0	3.095	0	1	1	Mixed
Part1	5	m3	m4	P	1.925	0.003	0.857	0.686	1	0	Mixed
Part1	5	m3	m5	P	1.864	0.004	0.945	0.553	1	0	Mixed
Part1	5	m4	m5	P	1.486	0.05	2.934	0	0	1	Mixed
Part3	5	m1	m2	P	1.257	0.17	1.58	0.029	0	1	Mixed
Part3	5	m1	m3	P	1.735	0.011	1.773	0.009	1	1	Mixed
Part3	5	m1	m4	P	1.631	0.021	1.223	0.199	1	0	Mixed
Part3	5	m1	m5	P	1.106	0.324	1.783	0.008	0	1	Mixed
Part3	5	m2	m3	P	1.67	0.017	0.882	0.649	1	0	Mixed

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part3	5	m2	m4	P	1.095	0.338	1.279	0.153	0	0	Mixed
Part3	5	m2	m5	P	1.149	0.273	1.022	0.437	0	0	Mixed
Part3	5	m3	m4	P	0.998	0.473	1.45	0.063	0	0	Mixed
Part3	5	m3	m5	P	1.08	0.358	1.57	0.031	0	1	Mixed
Part3	5	m4	m5	P	1.62	0.023	1.055	0.391	1	0	Mixed
Part1	7	m1	m2	H	0.511	0.987	0.71	0.875	0	0	m5
Part1	7	m1	m3	H	0.983	0.493	0.88	0.654	0	0	m5
Part1	7	m1	m4	H	0.512	0.987	1.603	0.022	0	1	m5
Part1	7	m1	m5	H	0.598	0.958	1.745	0.009	0	1	m5
Part1	7	m2	m3	H	0.884	0.647	1.266	0.157	0	0	m5
Part1	7	m2	m4	H	1.142	0.276	1.306	0.128	0	0	m5
Part1	7	m2	m5	H	0.504	0.988	0.925	0.583	0	0	m5
Part1	7	m3	m4	H	0.945	0.552	1.408	0.074	0	0	m5
Part1	7	m3	m5	H	0.785	0.789	1.565	0.029	0	1	m5
Part1	7	m4	m5	H	1.124	0.297	0.934	0.569	0	0	m5
Part2	7	m1	m2	P	1.34	0.108	0.876	0.659	0	0	Mixed
Part2	7	m1	m3	P	0.866	0.674	0.786	0.788	0	0	Mixed
Part2	7	m1	m4	P	0.684	0.899	0.899	0.624	0	0	Mixed
Part2	7	m1	m5	P	1.307	0.128	1.113	0.312	0	0	Mixed
Part2	7	m2	m3	P	1.429	0.066	0.945	0.552	0	0	Mixed
Part2	7	m2	m4	P	1.084	0.349	0.644	0.93	0	0	Mixed
Part2	7	m2	m5	P	0.907	0.611	1.007	0.457	0	0	Mixed
Part2	7	m3	m4	P	1.28	0.147	0.978	0.501	0	0	Mixed
Part2	7	m3	m5	P	0.917	0.596	0.841	0.711	0	0	Mixed
Part2	7	m4	m5	P	1.414	0.072	1.147	0.271	0	0	Mixed
Part3	7	m1	m2	H	1.838	0.005	2.273	0	1	1	m5
Part3	7	m1	m3	H	1.022	0.436	1.828	0.005	0	1	m5
Part3	7	m1	m4	H	3.119	0	0.851	0.696	1	0	m5
Part3	7	m1	m5	H	1.155	0.263	1.293	0.139	0	0	m5
Part3	7	m2	m3	H	1.323	0.119	1.799	0.006	0	1	m5
Part3	7	m2	m4	H	1.877	0.004	0.975	0.507	1	0	m5

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part3	7	m2	m5	H	2.019	0.001	1.865	0.004	1	1	m5
Part3	7	m3	m4	H	0.851	0.696	0.409	0.998	0	0	m5
Part3	7	m3	m5	H	0.737	0.845	0.886	0.644	0	0	m5
Part3	7	m4	m5	H	2.197	0	2.223	0	1	1	m5
Part1	8	m1	m2	H	0.972	0.51	1.562	0.031	0	1	m5
Part1	8	m1	m3	H	1.873	0.004	1.96	0.002	1	1	m5
Part1	8	m1	m4	H	2.866	0	1.925	0.003	1	1	m5
Part1	8	m1	m5	H	3.205	0	6.885	0	1	1	m5
Part1	8	m2	m3	H	1.024	0.433	1.251	0.172	0	0	m5
Part1	8	m2	m4	H	1.519	0.04	1.124	0.3	1	0	m5
Part1	8	m2	m5	H	1.697	0.013	2.083	0.001	1	1	m5
Part1	8	m3	m4	H	5.8	0	2.186	0	1	1	m5
Part1	8	m3	m5	H	4.77	0	1.349	0.105	1	0	m5
Part1	8	m4	m5	H	3.88	0	1.792	0.007	1	1	m5
Part2	8	m1	m2	H	1.27	0.158	0.79	0.781	0	0	m5
Part2	8	m1	m3	H	0.694	0.888	1.547	0.034	0	1	m5
Part2	8	m1	m4	H	0.912	0.604	1.037	0.415	0	0	m5
Part2	8	m1	m5	H	0.809	0.755	0.748	0.832	0	0	m5
Part2	8	m2	m3	H	0.942	0.557	0.889	0.638	0	0	m5
Part2	8	m2	m4	H	1.186	0.232	1.065	0.376	0	0	m5
Part2	8	m2	m5	H	0.784	0.789	1.334	0.115	0	0	m5
Part2	8	m3	m4	H	1.203	0.216	0.593	0.959	0	0	m5
Part2	8	m3	m5	H	0.888	0.64	1.149	0.271	0	0	m5
Part2	8	m4	m5	H	0.685	0.897	0.907	0.611	0	0	m5
Part2	9	m1	m2	P	1.743	0.009	1.259	0.163	1	0	Mixed
Part2	9	m1	m3	P	0.947	0.549	0.777	0.798	0	0	Mixed
Part2	9	m1	m4	P	0.825	0.734	1.148	0.27	0	0	Mixed
Part2	9	m1	m5	P	1.003	0.463	1	0.468	0	0	Mixed
Part2	9	m2	m3	P	1.075	0.361	1.398	0.079	0	0	Mixed
Part2	9	m2	m4	P	0.891	0.636	1.304	0.131	0	0	Mixed
Part2	9	m2	m5	P	0.861	0.681	0.664	0.915	0	0	Mixed

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part2	9	m3	m4	P	1.198	0.217	0.947	0.549	0	0	Mixed
Part2	9	m3	m5	P	1.266	0.158	1.37	0.092	0	0	Mixed
Part2	9	m4	m5	P	1.101	0.327	0.861	0.681	0	0	Mixed
Part1	10	m1	m2	P	0.915	0.598	1.785	0.007	0	1	Mixed
Part1	10	m1	m3	P	0.513	0.986	0.958	0.531	0	0	Mixed
Part1	10	m1	m4	P	1.495	0.044	0.985	0.491	1	0	Mixed
Part1	10	m1	m5	P	2.446	0	1.255	0.165	1	0	Mixed
Part1	10	m2	m3	P	0.969	0.514	0.68	0.903	0	0	Mixed
Part1	10	m2	m4	P	0.643	0.931	0.717	0.868	0	0	Mixed
Part1	10	m2	m5	P	1.27	0.154	0.802	0.767	0	0	Mixed
Part1	10	m3	m4	P	0.957	0.533	0.724	0.861	0	0	Mixed
Part1	10	m3	m5	P	1.023	0.434	1.251	0.169	0	0	Mixed
Part1	10	m4	m5	P	1.149	0.268	1.001	0.467	0	0	Mixed
Part1	11	m1	m2	H	1.254	0.168	3.225	0	0	1	m5
Part1	11	m1	m3	H	1.44	0.062	3.392	0	0	1	m5
Part1	11	m1	m4	H	1.482	0.049	3.71	0	1	1	m5
Part1	11	m1	m5	H	1.471	0.052	3.057	0	0	1	m5
Part1	11	m2	m3	H	1.845	0.004	5.165	0	1	1	m5
Part1	11	m2	m4	H	1.065	0.375	4.08	0	0	1	m5
Part1	11	m2	m5	H	2.708	0	3.838	0	1	1	m5
Part1	11	m3	m4	H	2.656	0	1.596	0.024	1	1	m5
Part1	11	m3	m5	H	4.195	0	1.725	0.01	1	1	m5
Part1	11	m4	m5	H	4.692	0	2.191	0	1	1	m5
Part2	11	m1	m2	P	1.705	0.012	1.074	0.362	1	0	Mixed
Part2	11	m1	m3	P	1.07	0.368	1.558	0.031	0	1	Mixed
Part2	11	m1	m4	P	1.154	0.263	1.064	0.375	0	0	Mixed
Part2	11	m1	m5	P	1.065	0.375	1.261	0.162	0	0	Mixed
Part2	11	m2	m3	P	0.718	0.866	0.934	0.569	0	0	Mixed
Part2	11	m2	m4	P	1.59	0.025	1.095	0.334	1	0	Mixed
Part2	11	m2	m5	P	1.138	0.282	0.624	0.943	0	0	Mixed
Part2	11	m3	m4	P	1.585	0.026	0.989	0.484	1	0	Mixed

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part2	11	m3	m5	P	0.923	0.586	1.235	0.183	0	0	Mixed
Part2	11	m4	m5	P	1.5	0.044	1.302	0.132	1	0	Mixed
Part1	13	m1	m2	P	1.059	0.383	2.552	0	0	1	Mixed
Part1	13	m1	m3	P	1.431	0.067	1.45	0.061	0	0	Mixed
Part1	13	m1	m4	P	2.208	0	1.439	0.064	1	0	Mixed
Part1	13	m1	m5	P	0.44	0.996	1.67	0.016	0	1	Mixed
Part1	13	m2	m3	P	2.127	0.001	0.468	0.993	1	0	Mixed
Part1	13	m2	m4	P	3.062	0	1.125	0.299	1	0	Mixed
Part1	13	m2	m5	P	1.724	0.011	1.471	0.054	1	0	Mixed
Part1	13	m3	m4	P	2.158	0	0.989	0.485	1	0	Mixed
Part1	13	m3	m5	P	0.844	0.706	1.996	0.002	0	1	Mixed
Part1	13	m4	m5	P	1.502	0.045	3.496	0	1	1	Mixed
Part2	13	m1	m2	H	1.275	0.148	1.207	0.205	0	0	m5
Part2	13	m1	m3	H	1.137	0.28	1.234	0.181	0	0	m5
Part2	13	m1	m4	H	1.51	0.039	0.942	0.557	1	0	m5
Part2	13	m1	m5	H	2.018	0.001	1.431	0.063	1	0	m5
Part2	13	m2	m3	H	0.768	0.811	0.768	0.812	0	0	m5
Part2	13	m2	m4	H	1.184	0.229	1.139	0.278	0	0	m5
Part2	13	m2	m5	H	0.597	0.959	1.159	0.256	0	0	m5
Part2	13	m3	m4	H	0.904	0.617	0.881	0.652	0	0	m5
Part2	13	m3	m5	H	0.93	0.576	0.957	0.533	0	0	m5
Part2	13	m4	m5	H	1.076	0.358	1.123	0.297	0	0	m5
Part2	14	m1	m2	P	0.716	0.869	0.863	0.679	0	0	Mixed
Part2	14	m1	m3	P	1.126	0.295	0.604	0.954	0	0	Mixed
Part2	14	m1	m4	P	1.651	0.016	1.014	0.447	1	0	Mixed
Part2	14	m1	m5	P	0.969	0.514	0.914	0.601	0	0	Mixed
Part2	14	m2	m3	P	0.957	0.534	0.773	0.804	0	0	Mixed
Part2	14	m2	m4	P	0.618	0.947	1.02	0.438	0	0	Mixed
Part2	14	m2	m5	P	0.522	0.984	1.223	0.193	0	0	Mixed
Part2	14	m3	m4	P	0.458	0.995	0.537	0.981	0	0	Mixed
Part2	14	m3	m5	P	0.717	0.868	0.52	0.985	0	0	Mixed

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part2	14	m4	m5	P	1.367	0.093	1.121	0.301	0	0	Mixed
Part1	16	m1	m2	P	1.097	0.334	1.579	0.029	0	1	Mixed
Part1	16	m1	m3	P	1.391	0.086	0.861	0.681	0	0	Mixed
Part1	16	m1	m4	P	0.732	0.85	1.019	0.442	0	0	Mixed
Part1	16	m1	m5	P	0.775	0.799	1.599	0.025	0	1	Mixed
Part1	16	m2	m3	P	0.828	0.729	0.712	0.871	0	0	Mixed
Part1	16	m2	m4	P	0.913	0.602	0.984	0.493	0	0	Mixed
Part1	16	m2	m5	P	1.082	0.353	1.509	0.044	0	1	Mixed
Part1	16	m3	m4	P	1.072	0.368	1.172	0.247	0	0	Mixed
Part1	16	m3	m5	P	1.302	0.135	1.238	0.185	0	0	Mixed
Part1	16	m4	m5	P	0.854	0.69	1.02	0.44	0	0	Mixed
Part1	17	m1	m2	P	1.506	0.043	1.327	0.118	1	0	Mixed
Part1	17	m1	m3	P	1.306	0.13	1.024	0.432	0	0	Mixed
Part1	17	m1	m4	P	0.886	0.644	1.117	0.308	0	0	Mixed
Part1	17	m1	m5	P	1.375	0.091	1.204	0.213	0	0	Mixed
Part1	17	m2	m3	P	1.23	0.189	1.423	0.07	0	0	Mixed
Part1	17	m2	m4	P	1.228	0.191	0.899	0.623	0	0	Mixed
Part1	17	m2	m5	P	1.205	0.212	0.881	0.652	0	0	Mixed
Part1	17	m3	m4	P	1.182	0.235	1.416	0.073	0	0	Mixed
Part1	17	m3	m5	P	1.172	0.245	0.982	0.495	0	0	Mixed
Part1	17	m4	m5	P	1.069	0.369	1.67	0.015	0	1	Mixed
Part2	17	m1	m2	H	1.602	0.024	1.299	0.137	1	0	m5
Part2	17	m1	m3	H	1.458	0.058	1.689	0.014	0	1	m5
Part2	17	m1	m4	H	3.051	0	1.291	0.142	1	0	m5
Part2	17	m1	m5	H	1.267	0.16	1.197	0.22	0	0	m5
Part2	17	m2	m3	H	1.698	0.013	2.454	0	1	1	m5
Part2	17	m2	m4	H	2.352	0	1.84	0.005	1	1	m5
Part2	17	m2	m5	H	1.486	0.05	2.228	0	1	1	m5
Part2	17	m3	m4	H	3.208	0	1.817	0.006	1	1	m5
Part2	17	m3	m5	H	1.78	0.008	2.474	0	1	1	m5
Part2	17	m4	m5	H	2.032	0.001	6.557	0	1	1	m5

Continued on next page

Table A.5 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part1	19	m1	m2	P	1.038	0.414	1.325	0.121	0	0	m1
Part1	19	m1	m3	P	0.966	0.52	1.311	0.131	0	0	m1
Part1	19	m1	m4	P	1.812	0.006	1.486	0.051	1	0	m1
Part1	19	m1	m5	P	1.719	0.012	1.361	0.101	1	0	m1
Part1	19	m2	m3	P	1.646	0.019	3.502	0	1	1	m1
Part1	19	m2	m4	P	1.802	0.007	2.029	0.001	1	1	m1
Part1	19	m2	m5	P	1.666	0.017	1.455	0.06	1	0	m1
Part1	19	m3	m4	P	1.799	0.007	1.766	0.009	1	1	m1
Part1	19	m3	m5	P	1.875	0.004	1.774	0.008	1	1	m1
Part1	19	m4	m5	P	1.205	0.214	1.621	0.022	0	1	m1
Part1	20	m1	m2	H	1.7	0.012	1.142	0.276	1	0	m5
Part1	20	m1	m3	H	1.074	0.362	1.676	0.014	0	1	m5
Part1	20	m1	m4	H	1.362	0.095	2.059	0.001	0	1	m5
Part1	20	m1	m5	H	1.559	0.029	1.947	0.002	1	1	m5
Part1	20	m2	m3	H	1.063	0.376	1.8	0.006	0	1	m5
Part1	20	m2	m4	H	1.308	0.126	2.111	0.001	0	1	m5
Part1	20	m2	m5	H	0.799	0.771	1.818	0.005	0	1	m5
Part1	20	m3	m4	H	1.479	0.048	1.61	0.021	1	1	m5
Part1	20	m3	m5	H	1.642	0.017	2.056	0.001	1	1	m5
Part1	20	m4	m5	H	1.862	0.004	1.377	0.087	1	0	m5

A.6 Table for Granger Causality test results in Borodin Quartet

The table below presents the results of the Granger Causality tests carried out on different parts of the Borodin Quartet. The “Part” column identifies the specific section of the Quintet being analysed. The ”File No.” column provides a sequential numbering of the analysed segments within each part. The “M1” and “M2” columns represent the first and second musicians in a dyadic pair being analysed. The “T” column signifies the texture of the musical piece, with

'P' representing 'Polyphonic' and 'H' denoting 'Homophonic'. The "F_M1_M2" and "F_M2_M1" columns provide the F values, which are statistical measures indicating the strength of causality from musician 1 to musician 2, and vice versa. The "p_M1_M2" and "p_M2_M1" columns present the corresponding p-values, which are probabilities used to determine the significance of the observed F values. The columns "p_M1_M2 (B)" and "p_M2_M1 (B)" hold binary values (1 or 0) indicating whether the pairs Granger causes each other or not, with 1 implying causality and 0 suggesting no causality. The "M_I" column signifies the main instrument involved in the Granger Causality analysis, providing additional details about the musicians' dyadic pair and the kind of instruments they played in each analysed segment. If a combination of instruments was involved, it is represented as "Mixed".

Table A.6: Detailed Granger Causality data for Borodin Concert Parts.

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part1	1	m1	m2	P	1.185	0.231	1.311	0.127	0	0	Mixed
Part1	1	m1	m3	P	0.622	0.944	0.692	0.891	0	0	Mixed
Part1	1	m1	m4	P	0.913	0.602	1.063	0.378	0	0	Mixed
Part1	1	m2	m3	P	1.035	0.418	1.041	0.409	0	0	Mixed
Part1	1	m2	m4	P	1.476	0.051	0.885	0.646	0	0	Mixed
Part1	1	m3	m4	P	0.588	0.962	1.181	0.235	0	0	Mixed
Part1	2	m1	m2	P	1.259	0.165	1.152	0.267	0	0	Mixed
Part1	2	m1	m3	P	0.771	0.806	1.049	0.397	0	0	Mixed
Part1	2	m1	m4	P	1.239	0.182	0.825	0.733	0	0	Mixed
Part1	2	m2	m3	P	1.104	0.324	0.814	0.749	0	0	Mixed
Part1	2	m2	m4	P	0.805	0.761	0.839	0.714	0	0	Mixed
Part1	2	m3	m4	P	1.36	0.099	1.437	0.065	0	0	Mixed
Part2	1	m1	m2	P	0.712	0.869	0.827	0.729	0	0	Mixed
Part2	1	m1	m3	P	0.76	0.817	0.702	0.879	0	0	Mixed
Part2	1	m1	m4	P	1.761	0.009	1.486	0.052	1	0	Mixed

Continued on next page

Table A.6 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part2	1	m2	m3	P	0.767	0.808	0.988	0.488	0	0	Mixed
Part2	1	m2	m4	P	1.456	0.061	0.827	0.729	0	0	Mixed
Part2	1	m3	m4	P	1.037	0.417	0.992	0.482	0	0	Mixed
Part2	2	m1	m2	H	1.263	0.163	1.151	0.269	0	0	m1
Part2	2	m1	m3	H	0.682	0.899	1.034	0.42	0	0	m1
Part2	2	m1	m4	H	0.824	0.735	1.048	0.4	0	0	m1
Part2	2	m2	m3	H	0.869	0.669	0.765	0.812	0	0	m1
Part2	2	m2	m4	H	0.888	0.639	1.18	0.238	0	0	m1
Part2	2	m3	m4	H	1.906	0.003	1.36	0.1	1	0	m1
Part2	3	m1	m2	H	1.354	0.103	1.239	0.182	0	0	m1
Part2	3	m1	m3	H	1.377	0.091	0.993	0.479	0	0	m1
Part2	3	m1	m4	H	1.301	0.135	1.17	0.249	0	0	m1
Part2	3	m2	m3	H	1.199	0.219	1.273	0.155	0	0	m1
Part2	3	m2	m4	H	0.934	0.569	1.183	0.235	0	0	m1
Part2	3	m3	m4	H	1.062	0.38	1.238	0.184	0	0	m1
Part2	4	m1	m2	P	0.573	0.969	1.587	0.025	0	1	Mixed
Part2	4	m1	m3	P	1.151	0.266	1.307	0.127	0	0	Mixed
Part2	4	m1	m4	P	0.837	0.718	0.932	0.572	0	0	Mixed
Part2	4	m2	m3	P	0.941	0.558	0.787	0.787	0	0	Mixed
Part2	4	m2	m4	P	1.498	0.043	1.465	0.053	1	0	Mixed
Part2	4	m3	m4	P	1.139	0.28	0.935	0.568	0	0	Mixed
Part2	5	m1	m2	P	1.024	0.433	1.192	0.223	0	0	Mixed
Part2	5	m1	m3	P	0.988	0.486	0.855	0.691	0	0	Mixed
Part2	5	m1	m4	P	0.936	0.567	0.61	0.951	0	0	Mixed
Part2	5	m2	m3	P	1.005	0.461	0.748	0.834	0	0	Mixed
Part2	5	m2	m4	P	0.871	0.666	1.351	0.102	0	0	Mixed
Part2	5	m3	m4	P	1.305	0.131	1.064	0.377	0	0	Mixed
Part2	6	m1	m2	H	1.348	0.104	1.073	0.363	0	0	m1
Part2	6	m1	m3	H	0.869	0.669	1.218	0.198	0	0	m1
Part2	6	m1	m4	H	1.548	0.033	1.238	0.181	1	0	m1
Part2	6	m2	m3	H	1.094	0.335	0.69	0.893	0	0	m1

Continued on next page

Table A.6 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part2	6	m2	m4	H	0.875	0.66	0.841	0.711	0	0	m1
Part2	6	m3	m4	H	0.985	0.49	0.473	0.993	0	0	m1
Part2	7	m1	m2	P	1.149	0.271	1.436	0.065	0	0	Mixed
Part2	7	m1	m3	P	1.453	0.059	0.989	0.485	0	0	Mixed
Part2	7	m1	m4	P	0.659	0.918	1.201	0.216	0	0	Mixed
Part2	7	m2	m3	P	1.022	0.436	0.755	0.825	0	0	Mixed
Part2	7	m2	m4	P	1.316	0.124	0.838	0.715	0	0	Mixed
Part2	7	m3	m4	P	1.05	0.396	1.169	0.248	0	0	Mixed
Part2	8	m1	m2	P	1.481	0.05	0.93	0.576	1	0	Mixed
Part2	8	m1	m3	P	2.201	0	1.062	0.379	1	0	Mixed
Part2	8	m1	m4	P	1.019	0.44	1.154	0.264	0	0	Mixed
Part2	8	m2	m3	P	1.279	0.149	1.208	0.209	0	0	Mixed
Part2	8	m2	m4	P	0.984	0.492	1.274	0.153	0	0	Mixed
Part2	8	m3	m4	P	1.331	0.115	1.105	0.323	0	0	Mixed
Part2	9	m1	m2	P	0.878	0.654	1.258	0.17	0	0	Mixed
Part2	9	m1	m3	P	0.95	0.546	1.17	0.252	0	0	Mixed
Part2	9	m1	m4	P	1.269	0.162	0.887	0.641	0	0	Mixed
Part2	9	m2	m3	P	1.203	0.218	1.026	0.432	0	0	Mixed
Part2	9	m2	m4	P	1.187	0.234	2.016	0.002	0	1	Mixed
Part2	9	m3	m4	P	0.926	0.581	1.045	0.405	0	0	Mixed
Part3	1	m1	m2	P	1.313	0.124	1.072	0.365	0	0	Mixed
Part3	1	m1	m3	P	0.952	0.542	1.304	0.13	0	0	Mixed
Part3	1	m1	m4	P	1.37	0.092	1.008	0.457	0	0	Mixed
Part3	1	m2	m3	P	1.086	0.346	0.877	0.657	0	0	Mixed
Part3	1	m2	m4	P	1.421	0.069	1.065	0.374	0	0	Mixed
Part3	1	m3	m4	P	0.939	0.561	1.088	0.343	0	0	Mixed
Part3	2	m1	m2	H	4.991	0	2.711	0	1	1	m1
Part3	2	m1	m3	H	8.441	0	3.541	0	1	1	m1
Part3	2	m1	m4	H	2.302	0	2.76	0	1	1	m1
Part3	2	m2	m3	H	2.551	0	2.814	0	1	1	m1
Part3	2	m2	m4	H	2.545	0	5.839	0	1	1	m1

Continued on next page

Table A.6 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part3	2	m3	m4	H	2.003	0.002	3.693	0	1	1	m1
Part4	1	m1	m2	H	1.149	0.271	1.3	0.136	0	0	m1
Part4	1	m1	m3	H	1.027	0.429	0.876	0.658	0	0	m1
Part4	1	m1	m4	H	1.151	0.269	1.119	0.306	0	0	m1
Part4	1	m2	m3	H	1.587	0.027	1.38	0.089	1	0	m1
Part4	1	m2	m4	H	0.665	0.913	0.584	0.963	0	0	m1
Part4	1	m3	m4	H	1.705	0.013	1.798	0.007	1	1	m1
Part4	2	m1	m2	H	0.889	0.639	0.991	0.481	0	0	m1
Part4	2	m1	m3	H	1.046	0.401	0.713	0.872	0	0	m1
Part4	2	m1	m4	H	1.124	0.299	0.652	0.925	0	0	m1
Part4	2	m2	m3	H	1.403	0.077	1.224	0.193	0	0	m1
Part4	2	m2	m4	H	1.296	0.136	0.953	0.54	0	0	m1
Part4	2	m3	m4	H	0.796	0.774	0.692	0.892	0	0	m1
Part4	3	m1	m2	H	0.788	0.782	0.638	0.932	0	0	m1
Part4	3	m1	m3	H	0.643	0.928	0.995	0.478	0	0	m1
Part4	3	m1	m4	H	0.653	0.922	0.933	0.572	0	0	m1
Part4	3	m2	m3	H	0.733	0.848	1.668	0.017	0	1	m1
Part4	3	m2	m4	H	1.2	0.221	0.416	0.998	0	0	m1
Part4	3	m3	m4	H	1.675	0.016	0.331	1	1	0	m1
Part4	4	m1	m2	H	1.066	0.376	1.618	0.023	0	1	m3
Part4	4	m1	m3	H	1.321	0.124	0.938	0.564	0	0	m3
Part4	4	m1	m4	H	0.534	0.98	1.415	0.075	0	0	m3
Part4	4	m2	m3	H	1.345	0.11	1.465	0.057	0	0	m3
Part4	4	m2	m4	H	1.538	0.037	2.264	0	1	1	m3
Part4	4	m3	m4	H	1.531	0.039	0.904	0.615	1	0	m3
Part4	5	m1	m2	H	1.176	0.244	2.464	0	0	1	m2
Part4	5	m1	m3	H	1.561	0.033	2.487	0	1	1	m2
Part4	5	m1	m4	H	1.323	0.123	1.594	0.027	0	1	m2
Part4	5	m2	m3	H	2.035	0.001	1.803	0.007	1	1	m2
Part4	5	m2	m4	H	1.597	0.026	0.991	0.483	1	0	m2
Part4	5	m3	m4	H	2.387	0	0.771	0.804	1	0	m2

Continued on next page

Table A.6 – continued from previous page

Part	FileNo	M1	M2	T	$F_{M1,M2}$	$p_{M1,M2}$	$F_{M2,M1}$	$p_{M2,M1}$	$p_{M1,M2}$ (B)	$p_{M2,M1}$ (B)	M_I
Part4	6	m1	m2	H	0.714	0.87	0.848	0.701	0	0	m1
Part4	6	m1	m3	H	1.483	0.049	0.778	0.797	1	0	m1
Part4	6	m1	m4	H	1.172	0.245	1.161	0.256	0	0	m1
Part4	6	m2	m3	H	1.432	0.066	1.205	0.211	0	0	m1
Part4	6	m2	m4	H	1.139	0.281	0.983	0.493	0	0	m1
Part4	6	m3	m4	H	0.81	0.755	0.542	0.979	0	0	m1
Part4	7	m1	m2	H	1.189	0.231	1.686	0.015	0	1	Mixed
Part4	7	m1	m3	H	1.237	0.187	0.812	0.751	0	0	Mixed
Part4	7	m1	m4	H	0.631	0.937	0.938	0.563	0	0	Mixed
Part4	7	m2	m3	H	1.2	0.22	1.25	0.176	0	0	Mixed
Part4	7	m2	m4	H	0.655	0.92	1.413	0.077	0	0	Mixed
Part4	7	m3	m4	H	1.029	0.427	0.887	0.641	0	0	Mixed

References

[1] M. L. Knapp, J. A. Hall, and T. G. Horgan, *Nonverbal communication in human interaction*. Cengage Learning, 2013. 1

[2] P. E. Keller, "Joint action in music performance," in *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions*, IOS press, 2008. 1, 16, 29, 94

[3] A. Chang, H. E. Kragness, S. R. Livingstone, D. J. Bosnyak, and L. J. Trainor, "Body sway reflects joint emotional expression in music ensemble performance," *Scientific Reports*, vol. 9, no. 1, pp. 1–11, 2019. 1, 99

[4] M. C. Van Der Steen and P. E. Keller, "The adaptation and anticipation model (adam) of sensorimotor synchronization," *Frontiers in human neuroscience*, vol. 7, p. 253, 2013. 1, 30, 96

[5] D. Glowinski, M. Mancini, R. Cowie, A. Camurri, C. Chiorri, and C. Doherty, "The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves," *Frontiers in psychology*, vol. 4, p. 841, 2013. 1, 96

[6] A. D'Ausilio, L. Badino, Y. Li, S. Tokay, L. Craighero, R. Canto, Y. Aloimonos, and L. Fadiga, "Leadership in orchestra emerges from the causal relationships of movement kinematics," *PLoS one*, vol. 7, no. 5, p. e35757, 2012. 1, 14, 17, 24, 33, 34, 96, 98

[7] L. Badino, A. D'Ausilio, D. Glowinski, A. Camurri, and L. Fadiga, "Sensorimotor communication in professional quartets," *Neuropsychologia*, vol. 55, pp. 98–104, 2014. 1, 32, 96

[8] K. Jokinen, "Non-verbal signals for turn-taking and feedback.," in *LREC*, 2010. 2

[9] M. Raković, N. F. Duarte, J. Marques, A. Billard, and J. Santos-Victor, "The gaze dialogue model: Nonverbal communication in hhi and hri," *IEEE Transactions on Cybernetics*, 2022. 2

[10] M. H. Woolhouse, D. Tidhar, and I. Cross, "Effects on inter-personal memory of dancing in time with others," *Frontiers in Psychology*, vol. 7, 2016. 2, 3

[11] A. Gaggioli, A. Chirico, E. Mazzoni, L. Milani, and G. Riva, "Networked flow in musical bands," *Psychology of Music*, vol. 45, pp. 283 – 297, 2017. 2, 3

[12] A. D'Ausilio, G. Novembre, L. Fadiga, and P. E. Keller, "What can music tell us about social interaction?," *Trends in cognitive sciences*, vol. 19, no. 3, pp. 111–114, 2015. 3, 100

[13] S. R. Sabharwal, M. Varlet, M. Breaden, G. Volpe, A. Camurri, and P. E. Keller, "husync - a model and system for the measure of synchronization in small groups: A case study on musical joint action," *IEEE Access*, vol. 10, pp. 92357–92372, 2022. 7, 95, 99, 107

[14] S. R. Sabharwal, A. Musso, M. Breaden, E. Riccomagno, A. Camurri, and P. E. Keller, *Analyzing directionality of influence among ensemble musicians using Granger Causality*, p. 107–118. Kansei Engineering and Emotion Research (KEER), Sep 2022. 8, 48, 101, 107

[15] S. R. Sabharwal, M. Breaden, G. Volpe, A. Camurri, and P. E. Keller, "Leadership dynamics in musical groups: Quantifying effects of musical structure on directionality of influence in concert performance videos," *PLOSOne*, 2023. Under Review. 8, 51, 108

[16] A. Bergsland and S. R. Sabharwal, "Examining the correlation between dance and electroacoustic music phrases: A pilot study," in *Proceedings of the 18th International Audio Mostly Conference*, AM '23, (New York, NY, USA), p. 34–37, Association for Computing Machinery, 2023. 8, 108

[17] U. Hess, "Nonverbal communication," *Encyclopedia of mental health*, vol. 2, no. 3, pp. 208–218, 2016. 11, 12

[18] C. Darwin, *The expression of the emotions in man and animals, 1872*. 1872. 11

[19] R. L. Birdwhistell, *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press, 2010. 11

[20] E. T. Hall, "A system for the notation of proxemic behavior," *American anthropologist*, vol. 65, no. 5, pp. 1003–1026, 1963. 11

[21] K. R. Scherer, "Personality inference from voice quality: The loud voice of extroversion," *European Journal of Social Psychology*, vol. 8, no. 4, pp. 467–487, 1978. 11

[22] H. Giles, "Accommodation theory: Optimal levels of convergence," *Language and social psychology*, pp. 45–65, 1979. 12

[23] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception–behavior link and social interaction.," *Journal of personality and social psychology*, vol. 76, no. 6, p. 893, 1999. 12

[24] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand, "The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry," *Journal of nonverbal behavior*, vol. 27, pp. 145–162, 2003. 12

[25] U. Hess and A. Fischer, "Emotional mimicry as social regulation," *Personality and social psychology review*, vol. 17, no. 2, pp. 142–157, 2013. 12

[26] H. Walter, "Social cognitive neuroscience of empathy: Concepts, circuits, and genes," *Emotion Review*, vol. 4, no. 1, pp. 9–17, 2012. 12

[27] M. Nusseck and M. M. Wanderley, "Music and motion——how music-related ancillary body movements contribute to the experience of music," *Music Perception*, vol. 26, no. 4, pp. 335–353, 2009. 13, 15

[28] J. W. Davidson, "Movement and collaboration in musical performance," *The Oxford handbook of music psychology*, pp. 364–376, 2009. 13, 14, 15, 95

[29] P. E. Keller, "Ensemble performance: Interpersonal alignment of musical expression," in *Expressiveness in music performance: Empirical approaches across styles and cultures*, pp. 260–282, Oxford University Press, 2014. 13, 14, 15, 16, 32, 94, 96, 98

[30] C. Palmer, "Music performance," *Annual review of psychology*, vol. 48, no. 1, pp. 115–138, 1997. 13

[31] R. Chaffin, A. F. Lemieux, and C. Chen, "It is different each time i play: Variability in highly prepared musical performance," *Music Perception*, vol. 24, no. 5, pp. 455–472, 2007. 13, 15

[32] A. Friberg and G. U. Battel, "Structural communication," *The science and psychology of music performance: Creative strategies for teaching and learning*, pp. 199–218, 2002. 13, 15

[33] A. Gabrielsson, "The performance of music," in *The psychology of music*, pp. 501–602, Elsevier, 1999. 13, 15

[34] B. H. Repp, "Musical synchronization," *Music, motor control, and the brain*, pp. 55–76, 2006. 13, 15

[35] P. E. Keller, G. Novembre, and M. J. Hove, "Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130394, 2014. 13, 14, 16, 17, 94, 97

[36] J. W. Davidson and J. M. Good, "Social and musical co-ordination between members of a string quartet: An exploratory study," *Psychology of music*, vol. 30, no. 2, pp. 186–201, 2002. 13, 15, 32

[37] R. Chaffin and T. Lisboa, "Practicing perfection: How concert soloists prepare for performance," *ICTUS-Periódico do PPGMUS-UFBA— ICTUS Music Journal*, vol. 9, no. 2, 2008. 14, 15

[38] J. Ginsborg, R. Chaffin, and G. Nicholson, "Shared performance cues in singing and conducting: A content analysis of talk during practice," *Psychology of Music*, vol. 34, no. 2, pp. 167–194, 2006. 14, 15

[39] A. C. Lehmann and K. A. Ericsson, "Preparation of a public piano performance: The relation between practice and performance," 1998. 14, 15

[40] P. E. Keller, "Attentional resource allocation in musical ensemble performance," *Psychology of Music*, vol. 29, no. 1, pp. 20–38, 2001. 14, 98

[41] F. A. Seddon and M. Biasutti, "Modes of communication between members of a string quartet," *Small group research*, vol. 40, no. 2, pp. 115–137, 2009. 14

[42] W. Goebl and C. Palmer, "Synchronization of timing and motion among performing musicians," *Music Perception*, vol. 26, no. 5, pp. 427–438, 2009. 14, 16, 17, 32, 97

[43] A. E. Walton, M. J. Richardson, P. Langland-Hassan, and A. Chemero, "Improvisation and the self-organization of multiple musical bodies," *Frontiers in psychology*, vol. 6, p. 313, 2015. 14, 32

[44] S. Dahl and A. Friberg, "Visual perception of expressiveness in musicians' body movements," *Music Perception*, vol. 24, no. 5, pp. 433–454, 2007. 14

[45] A. Camurri, G. Volpe, S. Piana, M. Mancini, R. Niewiadomski, N. Ferrari, and C. Canepa, "The dancer in the eye: towards a multi-layered computational framework of qualities in movement," in *Proceedings of the 3rd International Symposium on Movement and Computing*, pp. 1–7, 2016. 14, 91

[46] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of new music research*, vol. 33, no. 3, pp. 217–238, 2004. 14

[47] P. M. Hilt, L. Badino, A. D'Ausilio, G. Volpe, S. Tokay, L. Fadiga, and A. Camurri, "Multi-layer adaptation of group coordination in musical ensembles," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019. 14, 33, 35, 97, 98

[48] E. King and J. Ginsborg, “Gestures and glances: Interactions in ensemble rehearsal,” in *New perspectives on music and gesture*, pp. 203–228, Routledge, 2016. 14, 94

[49] E. King and C. Waddington, *Music and empathy.* Taylor & Francis, 2017. 14

[50] T. Lorenz, A. Weiss, and S. Hirche, “Synchrony and reciprocity: Key mechanisms for social companion robots in therapy and care,” *International Journal of Social Robotics*, vol. 8, pp. 125–143, 2016. 16, 17

[51] F. J. Bernieri and R. Rosenthal, “Interpersonal coordination: Behavior matching and interactional synchrony.,” 1991. 16

[52] K. R. Scherer and G. Ceschi, “Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport,” *Personality and Social Psychology Bulletin*, vol. 26, pp. 327 – 339, 2000. 16

[53] N. Sebanz, H. Bekkering, and G. Knoblich, “Joint action: bodies and minds moving together,” *Trends in Cognitive Sciences*, vol. 10, pp. 70–76, 2006. 16

[54] C. D. Frith, “The role of metacognition in human social interactions,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1599, pp. 2213–2223, 2012. 16

[55] M. Clayton, “What is entrainment? definition and applications in musical research,” 2012. 16

[56] K. L. Marsh, M. J. Richardson, and R. C. Schmidt, “Social connection through joint action and interpersonal coordination,” *Topics in Cognitive Science*, vol. 1, no. 2, pp. 320–339, 2009. 16, 17

[57] E. W. Large and M. R. Jones, "The dynamics of attending: How people track time-varying events.," *Psychological Review*, vol. 106, pp. 119–159, 1999. 16

[58] S. S. Wiltermuth and C. Heath, "Synchrony and cooperation," *Psychological science*, vol. 20, no. 1, pp. 1–5, 2009. 16, 18

[59] J. Buck, "Synchronous rhythmic flashing of fireflies. ii.," *The Quarterly review of biology*, vol. 63, no. 3, pp. 265–289, 1988. 16

[60] A. C. Homan, C. Buengeler, R. A. Eckhoff, W. P. van Ginkel, and S. C. Voelpel, "The interplay of diversity training and diversity beliefs on team creativity in nationality diverse teams.," *Journal of Applied Psychology*, vol. 100, no. 5, p. 1456, 2015. 17

[61] "Meet Pepper: The Robot Built for People — SoftBank Robotics America." 19

[62] E. Delaherche, M. Chetouani, F. Bigouret, J. Xavier, M. Plaza, and D. Cohen, "Assessment of the communicative and coordination skills of children with autism spectrum disorders and typically developing children using social signal processing," *Research in Autism Spectrum Disorders*, vol. 7, no. 6, pp. 741–756, 2013. 18, 20

[63] R. C. Schmidt and M. J. Richardson, "Dynamics of interpersonal coordination," in *Coordination: Neural, behavioral and social dynamics*, pp. 281–308, Springer, 2008. 18

[64] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 349–365, 2012. 18, 20, 21, 29, 30

[65] R. Feldman, "Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions," *Journal of Child psychology and Psychiatry*, vol. 48, no. 3-4, pp. 329–354, 2007. 18, 20

[66] J. Jaffe, B. Beebe, S. Feldstein, C. L. Crown, M. D. Jasnow, P. Rochat, and D. N. Stern, "Rhythms of dialogue in infancy: Coordinated timing in development," *Monographs of the society for research in child development*, pp. i–149, 2001. 18, 20

[67] M. J. Hove and J. L. Risen, "It's all in the timing: Interpersonal synchrony increases affiliation," *Social cognition*, vol. 27, no. 6, pp. 949–960, 2009. 18

[68] P. Valdesolo and D. DeSteno, "Synchrony and the social tuning of compassion.," *Emotion*, vol. 11, no. 2, p. 262, 2011. 18

[69] L. K. Miles, L. K. Nind, Z. Henderson, and C. N. Macrae, "Moving memories: Behavioral synchrony and memory for self and others," *Journal of Experimental Social Psychology*, vol. 46, no. 2, pp. 457–460, 2010. 18

[70] L. Noy, E. Dekel, and U. Alon, "The mirror game as a paradigm for studying the dynamics of two people improvising motion together," *Proceedings of the National Academy of Sciences*, vol. 108, no. 52, pp. 20947–20952, 2011. 20, 93

[71] F. Ramseyer and W. Tschacher, "Synchrony: A core concept for a constructivist approach to psychotherapy," *Constructivism in the human sciences*, vol. 11, no. 1-2, pp. 150–171, 2006. 20

[72] C. Trevarthen and S. Daniel, "Disorganized rhythm and synchrony: Early signs of autism and rett syndrome," *Brain and development*, vol. 27, pp. S25–S34, 2005. 20

[73] D. H. Abney, A. Paxton, R. Dale, and C. T. Kello, "Complexity matching in dyadic conversation.," *Journal of Experimental Psychology: General*, vol. 143, no. 6, p. 2304, 2014. 20

[74] F. Ramseyer and W. Tschacher, "Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome.," *Journal of consulting and clinical psychology*, vol. 79, no. 3, p. 284, 2011. 20, 31

[75] G. Varni, G. Volpe, and A. Camurri, "A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 576–590, 2010. 21

[76] G. Varni, A. Camurri, P. Coletta, and G. Volpe, "Emotional entrainment in music performance," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–5, IEEE, 2008. 21

[77] A. Camurri, G. Varni, and G. Volpe, "Measuring entrainment in small groups of musicians," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–4, IEEE, 2009. 21

[78] M. J. Richardson, K. L. Marsh, and R. Schmidt, "Effects of visual and verbal interaction on unintentional interpersonal coordination.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 1, p. 62, 2005. 22

[79] L. C. Freeman, D. Roeder, and R. R. Mulholland, "Centrality in social networks: Ii. experimental results," *Social networks*, vol. 2, no. 2, pp. 119–141, 1979. 23

[80] B. M. Bass and R. Bass, *The Bass handbook of leadership: Theory, research, and managerial applications*. Simon and Schuster, 2009. 23

[81] P. Gronn, "Distributed leadership as a unit of analysis," *The leadership quarterly*, vol. 13, no. 4, pp. 423–451, 2002. 23

[82] C. L. Pearce and J. A. Conger, *Shared leadership: Reframing the hows and whys of leadership*. Sage Publications, 2002. 23

[83] G. Yukl and R. Mahsud, "Why flexible and adaptive leadership is essential.," *Consulting Psychology Journal: practice and research*, vol. 62, no. 2, p. 81, 2010. 23

[84] A. K. Seth, "A matlab toolbox for granger causal connectivity analysis," *Journal of Neuroscience Methods*, vol. 186, no. 2, pp. 262–273, 2010. 23

[85] E. C. King, "The roles of student musicians in quartet rehearsals," *Psychology of music*, vol. 34, no. 2, pp. 262–282, 2006. 24

[86] T. A. Judge, J. E. Bono, R. Ilies, and M. W. Gerhardt, "Personality and leadership: a qualitative and quantitative review.," *Journal of applied psychology*, vol. 87, no. 4, p. 765, 2002. 24

[87] J. Antonakis, B. J. Avolio, and N. Sivasubramaniam, "Context and leadership: An examination of the nine-factor full-range leadership theory using the multifactor leadership questionnaire," *The leadership quarterly*, vol. 14, no. 3, pp. 261–295, 2003. 24

[88] C. Gockel and L. Werth, "Measuring and modeling shared leadership," *Journal of Personnel Psychology*, 2011. 24

[89] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pp. 3686–3693, 2014. 25

[90] Y. Yang and D. Ramanan, "Articulated human detection with flexible mixtures of parts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 12, pp. 2878–2890, 2012. 25

[91] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1653–1660, 2014. 25

[92] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 4724–4732, 2016. 25

[93] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pp. 483–499, Springer, 2016. 25

[94] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*, pp. 34–50, Springer, 2016. 26

[95] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017. 26

[96] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291–7299, 2017. 26, 27, 31

[97] S. Huang, M. Gong, and D. Tao, "A coarse-fine network for keypoint localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 3028–3037, 2017. 26

[98] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3d human pose estimation," in *Proceedings of the IEEE international conference on computer vision*, pp. 2640–2649, 2017. 26

[99] G. Moon, S.-I. Yu, H. Wen, T. Shiratori, and K. M. Lee, "Interhand2. 6m: A dataset and baseline for 3d interacting hand pose estimation from a single rgb image," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pp. 548–564, Springer, 2020. 26

[100] B. Wandt and B. Rosenhahn, "Repnet: Weakly supervised training of an adversarial reprojection network for 3d human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7782–7791, 2019. 26

[101] A. Lamas, S. Tabik, A. C. Montes, F. Pérez-Hernández, J. García, R. Olmos, and F. Herrera, "Human pose estimation for mitigating false negatives in weapon detection in video-surveillance," *Neurocomputing*, vol. 489, pp. 488–503, 2022. 27

[102] M. Einfalt, D. Zecha, and R. Lienhart, "Activity-conditioned continuous human pose estimation for performance analysis of athletes using the ex-

ample of swimming," in *2018 IEEE winter conference on applications of computer vision (WACV)*, pp. 446–455, IEEE, 2018. 27

[103] K. M. Kulkarni and S. Shenoy, "Table tennis stroke recognition using two-dimensional human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4576–4584, 2021. 27

[104] M. M. E. Yurtsever and S. Eken, "Babypose: Real-time decoding of baby's non-verbal communication using 2d video-based pose estimation," *IEEE Sensors Journal*, vol. 22, no. 14, pp. 13776–13784, 2022. 27

[105] F. Angelini, Z. Fu, Y. Long, L. Shao, and S. M. Naqvi, "Actionxpose: A novel 2d multi-view pose-based algorithm for real-time human action recognition," *arXiv preprint arXiv:1810.12126*, 2018. 27

[106] S. R. Fanello, I. Gori, G. Metta, and F. Odone, "Keep it simple and sparse: Real-time action recognition.," *Journal of Machine Learning Research*, vol. 14, 2013. 27

[107] D. Osokin, "Real-time 2d multi-person pose estimation on cpu: Lightweight openpose," *arXiv preprint arXiv:1811.12004*, 2018. 27

[108] S. K. Yadav, A. Singh, A. Gupta, and J. L. Raheja, "Real-time yoga recognition using deep learning," *Neural Computing and Applications*, vol. 31, pp. 9349–9361, 2019. 27

[109] F. M. Noori, B. Wallace, M. Z. Uddin, and J. Torresen, "A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network," in *Scandinavian conference on image analysis*, pp. 299–310, Springer, 2019. 27

[110] G. Goyal, F. Di Pietro, N. Carissimi, A. Glover, and C. Bartolozzi, "Moveenet: Online high-frequency human pose estimation with an event camera," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4023–4032, 2023. 28

[111] L. Song, G. Yu, J. Yuan, and Z. Liu, "Human pose estimation and its application to action recognition: A survey," *Journal of Visual Communication and Image Representation*, vol. 76, p. 103055, 2021. 28

[112] J. A. Hall, T. G. Horgan, and N. A. Murphy, "Nonverbal communication," *Annual review of psychology*, vol. 70, pp. 271–294, 2019. 29

[113] R. H. Humphrey, *Effective leadership: Theory, cases, and applications*. Sage Publications, 2013. 29

[114] N. E. Dunbar, J. K. Burgoon, and K. Fujiwara, "Automated methods to examine nonverbal synchrony in dyads," *Understanding Social Behavior in Dyadic and Small Group Interactions*, pp. 204–217, 2022. 29

[115] M. Chetouani, E. Delaherche, G. Dumas, and D. Cohen, "Interpersonal synchrony: From social perception to social interaction," *Social signal processing*, p. 202, 2017. 29, 30

[116] A. Chang, S. R. Livingstone, D. J. Bosnyak, and L. J. Trainor, "Body sway reflects leadership in joint music performance," *Proceedings of the National Academy of Sciences*, vol. 114, no. 21, pp. E4134–E4141, 2017. 30, 32, 33, 34, 52, 96, 98

[117] D. Newtson, "The dynamics of action and interaction.," 1993. 30

[118] D. Newtson, G. A. Engquist, and J. Bois, "The objective basis of behavior

units.," *Journal of Personality and social psychology*, vol. 35, no. 12, p. 847, 1977. 30

[119] D. Newtson, J. Hairfield, J. Bloomingdale, and S. Cutino, "The structure of action and interaction," *Social Cognition*, vol. 5, no. 3, pp. 191–237, 1987. 30

[120] R. C. Schmidt, S. Morr, P. Fitzpatrick, and M. J. Richardson, "Measuring the dynamics of interactional synchrony," *Journal of Nonverbal Behavior*, vol. 36, pp. 263–279, 2012. 30

[121] K. Tan, L. Xu, Y. Liu, and B. Luo, "Small group detection in crowds using interaction information," *IEICE TRANSACTIONS on Information and Systems*, vol. 100, no. 7, pp. 1542–1545, 2017. 30

[122] E. Novotny and G. Bente, "Identifying signatures of perceived interpersonal synchrony," *Journal of Nonverbal Behavior*, vol. 46, no. 4, pp. 485–517, 2022. 30

[123] M. Abtahi, *Brain-Body Sensory Fusion: Merging fNIRS/EEG Neuroimaging and Full-Body Motion Capture*. University of Rhode Island, 2018. 30

[124] G. Welch and E. Foxlin, "Motion tracking: No silver bullet, but a respectable arsenal," *IEEE Computer graphics and Applications*, vol. 22, no. 6, pp. 24–38, 2002. 31

[125] A. Paxton and R. Dale, "Frame-differencing methods for measuring bodily synchrony in conversation," *Behavior research methods*, vol. 45, no. 2, pp. 329–343, 2013. 31

[126] F. T. Ramseyer, "Motion energy analysis (mea): A primer on the assess-

ment of motion from video.," *Journal of counseling psychology*, vol. 67, no. 4, p. 536, 2020. 31

[127] A. Hadjakos, T. Großhauser, and W. Goebl, "Motion analysis of music ensembles with the kinect," in *Conference on New Interfaces for Musical Expression*, pp. 106–110, 2013. 31

[128] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, "Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 31

[129] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE winter conference on applications of computer vision (WACV)*, pp. 1–10, IEEE, 2016. 31

[130] K. Fujiwara and K. Yokomitsu, "Video-based tracking approach for nonverbal synchrony: a comparison of motion energy analysis and openpose," *Behavior Research Methods*, vol. 53, pp. 2700–2711, 2021. 31

[131] J. Laroche, A. Tomassini, G. Volpe, A. Camurri, L. Fadiga, and A. D'Ausilio, "Interpersonal sensorimotor communication shapes intrapersonal coordination in a musical ensemble," *Frontiers in Human Neuroscience*, vol. 16, p. 899676, 2022. 32

[132] J. MacRitchie and N. J. Bailey, "Efficient tracking of pianists' finger movements," *Journal of New Music Research*, vol. 42, no. 1, pp. 79–95, 2013. 32

[133] M. Ragert, T. Schroeder, and P. E. Keller, "Knowing too little or too much: the effects of familiarity with a co-performer's part on interpersonal

coordination in musical ensembles," *Frontiers in Psychology*, vol. 4, p. 368, 2013. 32, 75, 94, 98

[134] T. Eerola, G. Luck, P. Toiviainen, *et al.*, "An investigation of pre-schoolers' corporeal synchronization with music," in *Proceedings of the 9th international conference on music perception and cognition*, pp. 472–476, Alma Mater Studiorium University of Bologna, Bologna, Italy, 2006. 32

[135] R. Timmers, S. Endo, A. Bradbury, and A. M. Wing, "Synchronization and leadership in string quartet performance: a case study of auditory and visual cues," *Frontiers in Psychology*, vol. 5, p. 645, 2014. 32, 97

[136] A. M. Wing, S. Endo, A. Bradbury, and D. Vorberg, "Optimal feedback correction in string quartet synchronization," *Journal of The Royal Society Interface*, vol. 11, no. 93, p. 20131125, 2014. 32, 97

[137] L. Klein, E. A. Wood, D. Bosnyak, and L. J. Trainor, "Follow the sound of my violin: Granger causality reflects information flow in sound," *Frontiers in Human Neuroscience*, vol. 16, p. 982177, 2022. 32, 33

[138] L. Bishop, C. Cancino-Chacón, and W. Goebl, "Moving to communicate, moving to interact: Patterns of body motion in musical duo performance," *Music Perception: An Interdisciplinary Journal*, vol. 37, no. 1, pp. 1–25, 2019. 33, 94, 97, 101

[139] M. ROSENBLUM, P. TASS, J. Kurths, J. VOLKMANN, A. SCHNITZLER, and H.-J. FREUND, "Detection of phase locking from noisy data: application to magnetoencephalography," in *Chaos In Brain?*, pp. 34–51, World Scientific, 2000. 39

[140] F. Mormann, J. Fell, N. Axmacher, B. Weber, K. Lehnertz, C. E. Elger, and G. Fernández, "Phase/amplitude reset and theta–gamma interaction

in the human medial temporal lobe during a continuous word recognition memory task," *Hippocampus*, vol. 15, no. 7, pp. 890–900, 2005. 39

[141] S. Aydore, D. Pantazis, and R. M. Leahy, "A note on the phase locking value and its properties," *Neuroimage*, vol. 74, pp. 231–244, 2013. 39

[142] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964. 44

[143] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020. 44, 45

[144] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.*, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020. 45

[145] H. Luepsen, "Anova with binary variables: the f-test and some alternatives," *Communications in Statistics-Simulation and Computation*, vol. 52, no. 3, pp. 745–769, 2023. 49, 95

[146] Y. Antonacci, L. Minati, L. Faes, R. Pernice, G. Nollo, J. Toppi, A. Pietrabissa, and L. Astolfi, "Estimation of granger causality through artificial neural networks: applications to physiological systems and chaotic electronic oscillators," *PeerJ Computer Science*, vol. 7, p. e429, 2021. 49, 95

[147] J. R. Freeman, "Granger causality and the times series analysis of political relationships," *American Journal of Political Science*, pp. 327–358, 1983. 51

[148] Y. Zhou, Z. Kang, L. Zhang, and C. Spanos, "Causal analysis for nonstationary time series in sensor-rich smart buildings," in *2013 IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 593–598, IEEE, 2013. 51

[149] A. Zeileis, T. Hothorn, *et al.*, *Diagnostic checking in regression relationships*, vol. 2. na, 2002. 51

[150] C. D. Meals, "The question of lag: An exploration of the relationship between conductor gesture and sonic response in instrumental ensembles," *Frontiers in Psychology*, vol. 11, p. 573030, 2020. 52

[151] V. A. Brown, "An introduction to linear mixed-effects modeling in r," *Advances in Methods and Practices in Psychological Science*, vol. 4, no. 1, p. 2515245920960351, 2021. 53

[152] X. A. Harrison, L. Donaldson, M. E. Correa-Cano, J. Evans, D. N. Fisher, C. E. Goodwin, B. S. Robinson, D. J. Hodgson, and R. Inger, "A brief introduction to mixed effects modelling and multi-model inference in ecology," *PeerJ*, vol. 6, p. e4794, 2018. 53

[153] E. V. C. software], "Nijmegen: Max planck institute for psycholinguistics, the language archive," 2019. 62

[154] M. Clayton, K. Jakubowski, T. Eerola, P. E. Keller, A. Camurri, G. Volpe, and P. Alborno, "Interpersonal entrainment in music performance: theory, method, and model," *Music Perception: An Interdisciplinary Journal*, vol. 38, no. 2, pp. 136–194, 2020. 75, 95, 99

[155] T. Eerola, K. Jakubowski, N. Moran, P. E. Keller, and M. Clayton, "Shared periodic performer movements coordinate interactions in duo improvisations," *Royal Society Open Science*, vol. 5, no. 2, p. 171520, 2018. 75

[156] K. Jakubowski, T. Eerola, A. Blackwood Ximenes, W. K. Ma, M. Clayton, and P. E. Keller, "Multimodal perception of interpersonal synchrony: Evidence from global and continuous ratings of improvised musical duo performances.," *Psychomusicology: Music, Mind, and Brain*, vol. 30, no. 4, p. 159, 2020. 75

[157] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," in *Data analysis, machine learning and applications*, pp. 261–268, Springer, 2008. 75

[158] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, F. Scheipl, and G. Grothendieck, "Package 'lme4'," *Linear mixed-effects models using S4 classes. R package version*, vol. 1, no. 6, 2011. 75

[159] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. 75

[160] P. E. Keller and M. Appel, "Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles," *Music Perception*, vol. 28, no. 1, pp. 27–46, 2010. 75, 94, 97

[161] H. Luepsen, "Anova with binary variables: The f-test and some alternatives," *Communications in Statistics - Simulation and Computation*, vol. 52, pp. 745 – 769, 2021. 76

[162] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014. 80

[163] A. Pilotto, D. Sancarlo, F. Panza, F. Paris, G. D'Onofrio, L. Cascavilla, F. Addante, D. Seripa, V. Solfrizzi, B. Dallapiccola, *et al.*, "The multidimensional prognostic index (mpi), based on a comprehensive geriatric

assessment predicts short-and long-term mortality in hospitalized older patients with dementia," *Journal of Alzheimer's disease*, vol. 18, no. 1, pp. 191–199, 2009. 86

[164] K. Nymoen, *Methods and Technologies for Analysing Links Between Musical Sound and Body Motion.* Doctoral thesis, 2012. 90

[165] K. Nymoen, J. Torresen, R. I. Godøy, and A. R. Jensenius, "A statistical approach to analyzing sound tracings," pp. 120–145, Springer Berlin Heidelberg, 2012. 90

[166] G. Volpe, *Computational models of expressive gesture in multimedia systems.* PhD thesis, 01 2003. 90

[167] M. Zerr, C. Freihorst, H. Schütz, C. Sinke, A. Müller, S. Bleich, T. F. Münte, and G. R. Szycik, "Brief sensory training narrows the temporal binding window and enhances long-term multimodal speech perception," *Frontiers in Psychology*, vol. 10, 2019. 90

[168] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International conference on digital audio effects*, vol. 237, p. 244, Bordeaux, 2007. 90

[169] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Al-

gorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. 90

[170] R. Real and J. M. Vargas, "The probabilistic basis of jaccard's index of similarity," *Systematic Biology*, vol. 45, no. 3, pp. 380–385, 1996. 90

[171] R. T. Dean and F. Bailes, "Time series analysis as a method to examine acoustical influences on real-time perception of music," *Empirical Musicology Review*, vol. 5, pp. 152–175, 2010. 91

[172] L. Loke and T. Schiphorst, "The somatic turn in human-computer interaction," *Interactions*, vol. 25, no. 5, pp. 54–5863, 2018. 92

[173] L. Loke and T. Robertson, "Moving and making strange: An embodied approach to movement-based interaction design," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 20, no. 1, pp. 1–25, 2013. 92

[174] K. Hook, *Designing with the body: Somaesthetic interaction design.* MIT press, 2018. 92

[175] G. Novembre, M. Varlet, S. Muawiyath, C. J. Stevens, and P. E. Keller, "The e-music box: an empirical method for exploring the universal capacity for musical production and for social interaction through music," *Royal Society open science*, vol. 2, no. 11, p. 150286, 2015. 93, 97

[176] M. Varlet, S. Nozaradan, P. Nijhuis, and P. E. Keller, "Neural tracking and integration of 'self'and 'other'in improvised interpersonal coordination," *Neuroimage*, vol. 206, p. 116303, 2020. 93

[177] S. Kawase, "Gazing behavior and coordination during piano duo performance," *Attention, Perception, & Psychophysics*, vol. 76, no. 2, pp. 527–540, 2014. 94, 97

[178] L. Bishop and W. Goebl, "Communication for coordination: Gesture kinematics and conventionality affect synchronization success in piano duos," *Psychological Research*, vol. 82, no. 6, pp. 1177–1194, 2018. 94

[179] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin, "Cross-modal interactions in the perception of musical performance," *Cognition*, vol. 101, no. 1, pp. 80–113, 2006. 95

[180] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000. 95

[181] D. Glowinski, L. Badino, A. Ausilio, A. Camurri, and L. Fadiga, "Analysis of leadership in a string quartet," in *Third International Workshop on Social Behaviour in Music at ACM ICMI 2012*, pp. 763–774, Citeseer, 2012. 96

[182] L. Bishop, C. Cancino-Chacón, and W. Goebl, "Eye gaze as a means of giving and seeking information during musical interaction," *Consciousness and cognition*, vol. 68, pp. 73–96, 2019. 96, 97

[183] O. A. Heggli, J. Cabral, I. Konvalinka, P. Vuust, and M. L. Kringelbach, "A kuramoto model of self-other integration across interpersonal synchronization strategies," *PLoS computational biology*, vol. 15, no. 10, p. e1007422, 2019. 97

[184] I. Konvalinka, M. Bauer, C. Stahlhut, L. K. Hansen, A. Roepstorff, and C. D. Frith, "Frontal alpha oscillations distinguish leaders from followers: multivariate decoding of mutually interacting brains," *Neuroimage*, vol. 94, pp. 79–88, 2014. 97

[185] C. Alviar, R. Dale, A. Dewitt, and C. Kello, "Multimodal coordination of sound and movement in music and speech," *Discourse Processes*, vol. 57, no. 8, pp. 682–702, 2020. 97

[186] C. T. Kello, S. D. Bella, B. Médé, and R. Balasubramaniam, "Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes," *Journal of the Royal Society Interface*, vol. 14, no. 135, p. 20170231, 2017. 97

[187] L. Bishop, "Collaborative musical creativity: How ensembles coordinate spontaneity," *Frontiers in psychology*, vol. 9, p. 1285, 2018. 97

[188] G. Novembre, Z. Mitsopoulos, and P. E. Keller, "Empathic perspective taking promotes interpersonal coordination through music," *Scientific reports*, vol. 9, no. 1, p. 12255, 2019. 97

[189] C. Vesper, S. Butterfill, G. Knoblich, and N. Sebanz, "A minimal architecture for joint action," *Neural Networks*, vol. 23, no. 8-9, pp. 998–1003, 2010. 97

[190] J. MacRitchie, S. A. Herff, A. Procopio, and P. E. Keller, "Negotiating between individual and joint goals in ensemble musical performance," *Quarterly Journal of Experimental Psychology*, vol. 71, no. 7, pp. 1535–1551, 2018. 98, 104

[191] N. Pennill and R. Timmers, "Patterns of verbal interaction in newly formed music ensembles," *Frontiers in Psychology*, vol. 13, p. 987775, 2022. 98

[192] P. E. Keller and D. K. Burnham, "Musical meter in attention to multipart rhythm," *Music Perception*, vol. 22, no. 4, pp. 629–661, 2005. 98

[193] C. A. Coey, M. Varlet, and M. J. Richardson, "Coordination dynamics in a socially situated nervous system," *Frontiers in human neuroscience*, vol. 6, p. 164, 2012. 98

[194] B. B. Harry, D. S. Margulies, M. Falkiewicz, and P. E. Keller, "Brain networks for temporal adaptation, anticipation, and sensory-motor integration in rhythmic human behavior," *Neuropsychologia*, vol. 183, p. 108524, 2023. 98

[195] H. Lee and G. Orgs, "Experiencing art in social settings," *The Routledge International Handbook of Neuroaesthetics*, pp. 448–460, 2022. 99

[196] D. Moelants, M. Demey, M. Grachten, C.-F. Wu, and M. Leman, "The influence of an audience on performers: a comparison between rehearsal and concert using audio, video and movement data," *Journal of New Music Research*, vol. 41, no. 1, pp. 67–78, 2012. 99

[197] M. Wald-Fuhrmann, H. Egermann, A. Czepiel, K. O'Neill, C. Weining, D. Meier, W. Tschacher, F. Uhde, J. Toelle, and M. Tröndle, "Music listening in classical concerts: Theory, literature review, and research program," *Frontiers in psychology*, vol. 12, p. 638783, 2021. 99

[198] J. Davidson and M. C. Broughton, "Bodily mediated coordination, collaboration, and communication in music performance," in *The Oxford Handbook of Music Psychology*, Oxford University Press, 2016. 99, 102

[199] V. Sevdalis and P. E. Keller, "Perceiving performer identity and intended expression intensity in point-light displays of dance," *Psychological Research*, vol. 75, pp. 423–434, 2011. 99

[200] N. Jacoby, E. H. Margulis, M. Clayton, E. Hannon, H. Honing, J. Iversen, T. R. Klein, S. A. Mehr, L. Pearson, I. Peretz, *et al.*, "Cross-cultural work in music cognition: Challenges, insights, and recommendations," *Music Perception*, vol. 37, no. 3, pp. 185–195, 2020. 99

[201] D. Dotov, L. Delasanta, D. J. Cameron, E. W. Large, and L. Trainor, "Collective dynamics support group drumming, reduce variability, and stabilize tempo drift," *Elife*, vol. 11, p. e74816, 2022. 99

[202] A. P. Demos and C. Palmer, "Social and nonlinear dynamics unite: Musical group synchrony," *Trends in Cognitive Sciences*, 2022. 100

[203] O. A. Heggli, I. Konvalinka, M. L. Kringelbach, and P. Vuust, "A metastable attractor model of self–other integration (meamso) in rhythmic synchronization," *Philosophical Transactions of the Royal Society B*, vol. 376, no. 1835, p. 20200332, 2021. 100

[204] H. Liebermann-Jordanidis, G. Novembre, I. Koch, and P. E. Keller, "Simultaneous self-other integration and segregation support real-time interpersonal coordination in a musical joint action task," *Acta Psychologica*, vol. 218, p. 103348, 2021. 100

[205] G. Novembre, D. Sammler, and P. E. Keller, "Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action," *Neuropsychologia*, vol. 89, pp. 414–425, 2016. 100, 105

[206] G. Volpe, A. D'Ausilio, L. Badino, A. Camurri, and L. Fadiga, "Measuring social interaction in music ensembles," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1693, p. 20150377, 2016. 100

[207] J. Pearl *et al.*, "Models, reasoning and inference," *Cambridge, UK: CambridgeUniversityPress*, vol. 19, no. 2, p. 3, 2000. 101

[208] C. Palmer, F. Spidle, E. Koopmans, and P. Schubert, "Ears, heads, and eyes: when singers synchronise," *Quarterly Journal of Experimental Psychology*, vol. 72, no. 9, pp. 2272–2287, 2019. 101

[209] W. Forde Thompson and F. A. Russo, "Facing the music," *Psychological Science*, vol. 18, no. 9, pp. 756–757, 2007. 102

[210] S. R. Livingstone and C. Palmer, "Head movements encode emotions during speech and song.," *Emotion*, vol. 16, no. 3, p. 365, 2016. 102

[211] B. Buck, J. MacRitchie, and N. J. Bailey, "The interpretive shaping of embodied musical structure in piano performance," *Empirical Musicology Review*, pp. 92–119, 2013. 102

[212] J. W. Davidson, "Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies," *Psychology of Music*, vol. 40, no. 5, pp. 595–633, 2012. 102

[213] D. Glowinski, F. Dardard, G. Gnecco, S. Piana, and A. Camurri, "Expressive non-verbal interaction in a string quartet: an analysis through head movements," *Journal on Multimodal User Interfaces*, vol. 9, no. 1, pp. 55–68, 2015. 102

[214] M. Zhou, L. Stoffl, M. Mathis, and A. Mathis, "Rethinking pose estimation in crowds: overcoming the detection information-bottleneck and ambiguity," *arXiv preprint arXiv:2306.07879*, 2023. 103

[215] Z. Tang, R. Gu, and J.-N. Hwang, "Joint multi-view people tracking and pose estimation for 3d scene reconstruction," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2018. 104

[216] M.-H. Nguyen, C.-C. Hsiao, W.-H. Cheng, and C.-C. Huang, "Practical 3d human skeleton tracking based on multi-view and multi-kinect fusion," *Multimedia Systems*, vol. 28, no. 2, pp. 529–552, 2022. 104

[217] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," *ACM transactions on graphics (TOG)*, vol. 22, no. 3, pp. 569–577, 2003. 104

[218] P. E. Keller, G. Novembre, and J. Loehr, "Musical ensemble performance: representing self, other and joint action outcomes," *Shared representations: Sensorimotor foundations of social life*, p. 280, 2016. 105

[219] E. A. Wood, A. Chang, D. Bosnyak, L. Klein, E. Baraku, D. Dotov, and L. J. Trainor, "Creating a shared musical interpretation: Changes in coordination dynamics while learning unfamiliar music together," *Annals of the New York Academy of Sciences*, vol. 1516, no. 1, pp. 106–113, 2022. 105

[220] V. Müller, K.-R. P. Ohström, and U. Lindenberger, "Interactive brains, social minds: neural and physiological mechanisms of interpersonal action coordination," *Neuroscience & Biobehavioral Reviews*, vol. 128, pp. 661–677, 2021. 105

[221] B. Rimé and L. Schiaratura, "Gesture and speech in fundamentals of nonverbal behavior," 01 1991.

[222] E. Coronado, J. Villalobos, B. Bruno, and F. Mastrogiovanni, "Gesture-based Robot Control: Design Challenges and Evaluation with Humans," 05 2017.

[223] N. Jaouedi, N. Boujnah, O. Htiwich, and M. S. Bouhlel, "Human action recognition to human behavior analysis," in *2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, pp. 263–266, IEEE, 2016.

[224] B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE transactions on multimedia*, vol. 17, no. 7, pp. 1107–1119, 2015.

[225] G. Luck, S. Saarikallio, B. Burger, M. R. Thompson, and P. Toiviainen, "Effects of the big five and musical genre on music-induced movement," *Journal of Research in Personality*, vol. 44, no. 6, pp. 714–720, 2010.

[226] J. K. Vuoskoski, W. F. Thompson, D. McIlwain, and T. Eerola, "Who enjoys listening to sad music and why?," *Music Perception*, vol. 29, no. 3, pp. 311–317, 2011.

[227] M. M. Wanderley, "Quantitative analysis of non-obvious performer gestures," in *International Gesture Workshop*, pp. 241–253, Springer, 2001.

[228] L. Bishop and W. Goebl, "Beating time: How ensemble musicians' cueing gestures communicate beat position and tempo," *Psychology of Music*, vol. 46, no. 1, pp. 84–106, 2018.

[229] G. Luck and P. Toiviainen, "Ensemble musicians' synchronization with conductors' gestures: An automated feature-extraction analysis," *Music Perception*, vol. 24, no. 2, pp. 189–200, 2006.

[230] B. Burger, M. R. Thompson, G. Luck, S. H. Saarikallio, and P. Toiviainen, "Hunting for the beat in the body: on period and phase locking in music-induced movement," *Frontiers in human neuroscience*, vol. 8, p. 903, 2014.

[231] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in *Proceedings of the IEEE international conference on computer vision*, pp. 2334–2343, 2017.

[232] B. Schögler, "Studying temporal co-ordination in jazz duets," *Musicae scientiae*, vol. 3, no. 1_suppl, pp. 75–91, 1999.

[233] K. Yun, K. Watanabe, and S. Shimojo, "Interpersonal body and neural synchronization as a marker of implicit social interaction," *Scientific reports*, vol. 2, no. 1, pp. 1–8, 2012.

[234] A. C. Onslow, R. Bogacz, and M. W. Jones, "Quantifying phase–amplitude coupling in neuronal network oscillations," *Progress in biophysics and molecular biology*, vol. 105, no. 1-2, pp. 49–57, 2011.

[235] O. Jensen and J. E. Lisman, "An oscillatory short-term memory buffer model can account for data on the sternberg task," *Journal of Neuroscience*, vol. 18, no. 24, pp. 10688–10699, 1998.

[236] O. Jensen, "Maintenance of multiple working memory items by temporal segmentation," *Neuroscience*, vol. 139, no. 1, pp. 237–249, 2006.

[237] J. E. Lisman and O. Jensen, "The theta-gamma neural code," *Neuron*, vol. 77, no. 6, pp. 1002–1016, 2013.

[238] J. Vosskuhl, R. J. Huster, and C. S. Herrmann, "Increase in short-term memory capacity induced by down-regulating individual theta frequency via transcranial alternating current stimulation," *Frontiers in human neuroscience*, vol. 9, p. 257, 2015.

[239] G. Varni, M. Mancini, L. Fadiga, A. Camurri, and G. Volpe, "The change matters! measuring the effect of changing the leader in joint music performances," *IEEE Transactions on Affective Computing*, 2019.

[240] Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, and A.-L. Barabási, "The sound of many hands clapping," *Nature*, vol. 403, no. 6772, pp. 849–850, 2000.

[241] S. Kirschner and M. Tomasello, "Joint drumming: Social context facilitates synchronization in preschool children," *Journal of experimental child psychology*, vol. 102, no. 3, pp. 299–314, 2009.

[242] M. J. Richardson, K. L. Marsh, R. W. Isenhower, J. R. Goodman, and R. C. Schmidt, "Rocking together: Dynamics of intentional and unintentional interpersonal coordination," *Human movement science*, vol. 26, no. 6, pp. 867–891, 2007.

[243] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, "Multimodal analysis of expressive gesture in music and dance performances," in *International gesture workshop*, pp. 20–39, Springer, 2003.

[244] A. Jensenius, M. Wanderley, R. Godøy, and M. Leman, "Concepts and methods in research on music-related gestures," *Musical gestures: Sound, movement, and meaning*, pp. 12–35, 2010.

[245] J. MacRitchie, B. Buck, and N. J. Bailey, "Inferring musical structure through bodily gestures," *Musicae Scientiae*, vol. 17, no. 1, pp. 86–108, 2013.

[246] E. Goodman, "Ensemble performance," *Musical Performance*, 2002.

[247] S. D'Amario, H. Daffern, and F. Bailes, "Synchronization in singing duo performances: The roles of visual contact and leadership instruction," *Frontiers in Psychology*, vol. 9, p. 1208, 2018.

[248] D. Glowinski, F. Bracco, C. Chiorri, and D. Grandjean, "Music ensemble as a resilient system. managing the unexpected through group interaction," *Frontiers in Psychology*, vol. 7, p. 1548, 2016.

[249] P. Alborno, G. Volpe, M. Mancini, R. Niewiadomski, S. Piana, and A. Camurri, "The multi-event-class synchronization (mecs) algorithm," *arXiv preprint arXiv:1903.09530*, 2019.

[250] M. Tal-Shmotkin and A. Gilboa, "Do behaviors of string quartet ensembles represent self-managed teams?," *Team Performance Management: An International Journal*, 2013.

[251] J. Decety and W. Ickes, *The social neuroscience of empathy.* MIT press, 2011.

[252] N. Latif, A. V. Barbosa, E. Vatiokiotis-Bateson, M. S. Castelhano, and K. Munhall, "Movement coordination during conversation," *PLoS one*, vol. 9, no. 8, p. e105036, 2014.

[253] T. Yoshida, S. Takeda, and S. Yamamoto, "The application of entrainment to musical ensembles," in *II International Conference on Music and Artificial Intelligence (ICMAI), Edinburgh, Scotland*, Citeseer, 2002.

[254] S. Cole and B. Voytek, "Cycle-by-cycle analysis of neural oscillations," *Journal of neurophysiology*, vol. 122, no. 2, pp. 849–861, 2019.

[255] J. M. Hurtado, L. L. Rubchinsky, and K. A. Sigvardt, "Statistical method for detection of phase-locking episodes in neural oscillations," *Journal of neurophysiology*, vol. 91, no. 4, pp. 1883–1898, 2004.

[256] J. project (2021), "jamovi (version 1.6) [computer software]. the jamovi project. retrieved from https://www.jamovi.org."

[257] C. Nagaoka and M. Komori, "Body movement synchrony in psychotherapeutic counseling: A study using the video-based quantification method,"

IEICE transactions on information and systems, vol. 91, no. 6, pp. 1634–1640, 2008.

[258] M. Komori, "A video-based quantification method of body movement synchrony: An application for dialogue in counseling," *Jpn. J. Interpersonal and Social Psychology*, vol. 7, pp. 41–48, 2007.

[259] T. Yokozuka, E. Ono, Y. Inoue, K.-I. Ogawa, and Y. Miyake, "The relationship between head motion synchronization and empathy in unidirectional face-to-face communication," *Frontiers in psychology*, p. 1622, 2018.

[260] K. Jakubowski, T. Eerola, P. Alborno, G. Volpe, A. Camurri, and M. Clayton, "Extracting coarse body movements from video in music performance: a comparison of automated computer vision techniques with motion capture data," *Frontiers in Digital Humanities*, vol. 4, p. 9, 2017.

[261] P. Toiviainen, V. Alluri, E. Brattico, M. Wallentin, and P. Vuust, "Capturing the musical brain with lasso: Dynamic decoding of musical features from fmri data," *Neuroimage*, vol. 88, pp. 170–180, 2014.

[262] C. Cornejo, Z. Cuadros, R. Morales, and J. Paredes, "Interpersonal coordination: methods, achievements, and challenges," *Frontiers in psychology*, vol. 8, p. 1685, 2017.

[263] C. Wöllner and P. E. Keller, "Music with others: Ensembles, conductors, and interpersonal coordination," in *The Routledge companion to music cognition*, pp. 313–324, Routledge, 2017.

[264] A. Lillywhite, D. Glowinski, A. Camurri, and F. Pollick, "Using fmri and intersubject correlation to explore brain activity during audiovisual observation of a string quartet," *Front. Hum. Neurosci.*, 2015.

[265] A. Williamon and J. W. Davidson, "Exploring co-performer communication," *Musicae Scientiae*, vol. 6, no. 1, pp. 53–72, 2002.

[266] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[267] J. W. Davidson, "Visual perception of performance manner in the movements of solo musicians," *Psychology of music*, vol. 21, no. 2, pp. 103–113, 1993.

[268] S. D'Amario, H. Daffern, and F. Bailes, "A new method of onset and offset detection in ensemble singing," *Logopedics Phoniatrics Vocology*, vol. 44, no. 4, pp. 143–158, 2019.

[269] M. P. Silvernagel, A. S. Ling, and P. Nuyujukian, "A markerless platform for ambulatory systems neuroscience," *Science Robotics*, vol. 6, 2021.

[270] J. Marina-Miranda and V. J. Traver, "Head and eye egocentric gesture recognition for human-robot interaction using eyewear cameras," *IEEE Robotics and Automation Letters*, vol. 7, pp. 7067–7074, 2022.

[271] J. G. Han, N. Campbell, K. Jokinen, and G. Wilcock, "Investigating the use of non-verbal cues in human-robot interaction with a nao robot," *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 679–683, 2012.

[272] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1653–1660, 2014.

[273] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, "Towards accurate multi-person pose estimation in the wild," 2017.

[274] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, pp. 1297–1304, 2011.

[275] B. Caramiaux, A. Altavilla, S. G. Pobiner, and A. Tanaka, "Form follows sound: designing interactions from sonic memories," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3943–3952, 2015.

[276] J. B. Bavelas, A. Black, C. R. Lemery, and J. Mullett, "" i show how you feel": Motor mimicry as a communicative act.," *Journal of personality and social psychology*, vol. 50, no. 2, p. 322, 1986.

[277] A. I. Goldman and C. S. Sripada, "Simulationist models of face-based emotion recognition," *Cognition*, vol. 94, no. 3, pp. 193–213, 2005.

[278] T. Lipps, *Ästhetik*. Teubner, 1907.