Interpersonal Synchronization in Ensemble Singing: The Roles of Visual Contact and Leadership, and Evolution across Rehearsals

Sara D’Amario

PhD

University of York

Electronic Engineering

November 2018

Abstract

Interpersonal synchronization between musicians in Western ensembles is a fundamental performance parameter, contributing to the expressiveness of ensemble performances. Synchronization might be affected by the visual contact between musicians, leadership, and rehearsals, although the nature of these relationships has not been fully investigated. This thesis centres on the synchronization between singers in *a cappella* singing ensembles, in relation to the roles of visual cues and leadership instruction in 12 duos, and the evolution of synchronization and leader-follower relationships emerging spontaneously across five rehearsals in a newly formed quintet. In addition, the developmental aspects of synchronization are investigated in parallel to tuning and verbal interactions, to contextualise synchronization within the wider scope of expressive performance behaviours.

Three empirical investigations were conducted to study synchronization in singing ensembles, through a novel algorithm developed for this research, based on the application of electrolaryngography and acoustic analysis. Findings indicate that synchronisation is a complex issue in terms of performance and perception. Synchronization was better with visual contact between singers than without in singing duos, and improved across rehearsals in the quintet depending on the piece performed. Leadership instruction did not affect precision or consistency of synchronization in singing duos; however, when the upper voice was instructed to lead, the designated leader preceded the co-performer. Leadership changed across rehearsals, becoming equally distributed in the last rehearsal. Differences in the precision of synchronization related to altered visual contact were reflected in the perception of synchronization irrespective of the listeners' music expertise, but the smaller asynchrony patterns measured across rehearsals were not. Synchronization in the quintet was not the result of rehearsal strategies targeted for the purpose of synchronization during rehearsal, but was paired with a tendency to tune horizontally towards equal temperament (ET), and to ET and just intonation in the vertical tuning of third intervals.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest appreciation and gratitude to my supervisors, Helena Daffern and Freya Bailes, for their patient guidance and mentorship throughout this research. They definitely provided me with the tools and encouragement needed for the successful completion of this thesis.

I am grateful to Renee Timmers for her advice on this research, and to David Howard for his pastoral care on the study of tuning. A heartfelt thanks to Johan Sundberg for the fruitful discussions on the singing voice, and to Benoît Bardy, who broadened my horizons on interpersonal synchronization. I am truly indebted and thankful to Sten Ternström, who provided insight, expertise, motivation, encouragement, and a genuine friendship that greatly supported my research visit at the Royal Institute of Technology, KTH, in Stockholm and my PhD.

I would also like to thank the White Rose College of Arts and Humanities for funding this PhD. Thanks to all the University of York Audio Lab members and my colleagues, Nicola Pennill and Ryan Kirkbride, for their friendship, support, and advice. Thanks to Francis Duah of the Math and Stats Centre at the University of York for the extremely knowledgeable support with the multilevel linear modelling.

Thanks to my parents, and to my brother and sister, for always being there when I needed their love, and help with my children. Thanks to my nieces, Martina, Aurora, and Camilla to have brought so much joy on my life. Thanks to my grandparents for instilling perseverance in me.

An immense thanks to my gorgeous babies, Giulia and Laura Audrey, for filling up all days of my life with love, smiles, and so much needed cuddling. And last, but definitely by no means least, thanks to my lovely husband Giorgio, because I wouldn't have completed this research without him. He is behind every word of this thesis; every idea and every concept of this project. He is simply my better half.

Declaration

I hereby declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this or any other University. All sources are explicitly stated and referenced.

I also declare that parts of this thesis have been published in peer-reviewed academic journals, as follows:

- **D'Amario, Sara**, David Howard, Helena Daffern, and Nicola Pennill. (2018). "A Longitudinal Study of Intonation in an a Cappella Singing Quintet*." Journal of Voice.* https://doi.org/10.1016/j.jvoice.2018.07.015.
- **D'Amario, Sara**, Helena Daffern, and Freya Bailes. (2018). "A Longitudinal Study Investigating Synchronization in Singing Quintet Performances." *Journal of Voice.* https://doi.org/10.1016/j.jvoice.2018.07.015.
- **D'Amario, Sara**, Helena Daffern, and Freya Bailes. (2018). "Synchronization in Singing Duo Performances: The Roles of Visual Contact and Leadership Instruction." *Frontiers in Psychology* 9 (July): 1208*.* https://doi.org/10.3389/fpsyg.2018.01208.
- **D'Amario, Sara**, Helena Daffern, and Freya Bailes. (2018). "A New Method for Onset and Offset Detection in Ensemble Singing." *Logopedics, Phoniatrics and Vocology.* https://doi.org/10.1080/14015439.2018.1452977.
- **D'Amario**, **Sara**, and Helena Daffern. (2017). "Using Electrolaryngography and Electroglottography to Assess the Singing Voice: A Systematic Review." *Psychomusicology: Music, Mind, and Brain, 26*(4), 229-243. http://dx.doi.org/10.1037/pmu0000184.

Furthermore, part of this thesis is currently under review for publication in peer-reviewed journals, as follows:

- **D'Amario, Sara**, Helena Daffern, and Freya Bailes. (In preparation). "Perception of Synchronization in Singing Ensembles". *PlosOne*.
- **D'Amario, Sara**, and Helena Daffern. (In preparation). "Understanding Singing Performance through the Lens of Empirical Acoustics Research." *AI & SOCIETY Journal of Knowledge, Culture and Communication.*

In addition, other parts of this thesis have been presented at national and international conferences, as follows:

- **D’Amario, Sara**, Helena Daffern, and Freya Bailes. (2018). “Do Performed Asynchronies Bear a Relationship to the Synchrony that Listeners with a Variety of Levels of Musical Experience Can Perceive?” *Paper presented at the 15th International Conference on Music Perception and Cognition - 10th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23rd – July 28th 2018*, Graz, Austria.
- **D’Amario, Sara**, Helena Daffern, and Freya Bailes. (2018). “A Longitudinal Study Investigating Synchronization in Singing Quintet Performances.” *Paper presented at The Voice Foundation’s 47th Annual Symposium: Care of the Professional Voice, May 30th – June 3rd 2018,* Philadelphia, USA.
- **D’Amario, Sara**, David Howard, and Helena Daffern. (2018). “A Longitudinal Study of Intonation in a Cappella Quintet Singing.” *Paper presented at The Voice Foundation’s 47th Annual Symposium: Care of the Professional Voice, May 30th – June 3rd 2018*, Philadelphia, USA.
- **D’Amario, Sara,** Helena Daffern, and Freya Bailes. (2018). “A Study of the Effects of Music Training and Performance Experience on Perceptions of Singing Ensemble Synchronization.” *Together in Music (TiM), April 12th- 14th 2018*, York, UK.
- Pennill, Nicola, and **Sara D’Amario.** (2018). “Development of Rehearsal Interactions in a Newly-Formed Vocal Quintet.” *Together in Music (TiM), April 12th- 14th 2018*, York, UK.
- **D’Amario, Sara**, Helena Daffern, and Freya Bailes. (2017). “Temporal Synchronization in Singing Duet Performances.” *Paper presented at the Pan-European Voice Conference (PEVOC), August 30th – September 1st 2017*, Ghent, Belgium.
- **D’Amario, Sara,** and Helena Daffern. (2017). “Interpersonal Interaction of Portamento in Singing Ensemble Performances.*” Poster session presented at the European Society for Cognitive Sciences Of Music (ESCOM), July 31st – August 4th 2017*, Ghent, Belgium.
- **D’Amario, Sara**, Helena Daffern, and Freya Bailes. (2017). “Synchronization in Singing Duet Performances: The Roles of Leadership and Visual Contact.” *Paper presented at*

The Voice Foundation's 46th Annual Symposium: Care of the Professional Voice, May 31st – June 4th 2017, Philadelphia, USA.

- **D'Amario, Sara**, Ryan Kirkbride, and Nicola Pennill. (2016). "A Preliminary Observational Study of Vocal Quartets in a Masterclass Setting." *Paper presented at White Rose College of the Arts and Humanities Conference, October 6th 2016*, York, UK.
- **D'Amario, Sara**, and Helena Daffern. (2016). "Does a Masterclass Improve Group Cohesion? A Pilot Study of Temporal Synchronization between Singers during Vocal Quartet Performances." *Paper presented at the Music Education and Music Psychology Student Conference, June 23rd 2016,* Department of Music, University of York, York, UK.
- **D'Amario, Sara**, and Helena Daffern. (2016). "Electrolaryngography and Electroglottography in the Assessment of Singing Voice: A Systematic Literature Review." *Poster session presented at 9th International Conference of Students of Systematic Musicology (SYSMUS), June 8th – 10th 2016*, Department of Music, University of Jyväskylä, Jyväskylä, Finland.
- **D'Amario, Sara**, and Helena Daffern. (2016). "Understanding Singing Performances through the Lens of Empirical Acoustics Research." *Poster session presented at the Performing Knowledge Conference, April 25th – 26th 2016*, Emmanuel College, Cambridge, UK.

1 Introduction

> *We try ... to avoid impositions (of rhythm and tempo). If one player takes a little musical liberty, the quartet goes along with him. We allow each other freedom—but there's a natural give and take ... A moment of* ritardando *or* rubato *should not sound contrived (by being planned); it should be allowed to happen naturally.*
>
> David Soyer, cellist in the Guarnier Quartet,
> Blum (1987, 16)

How is it possible to maintain a *natural* music performance if *one player takes a little musical liberty* during ensemble playing? A complex set of factors fosters cohesive and natural synchronization between musicians during ensemble performances. But little is known about the roles of visual contact between musicians, leader-follower relationships, and rehearsal on interpersonal synchronization during ensemble playing, and particularly singing.

1.1 The Notion of Synchronization

Synchronization, broadly defined, refers to the temporal coordination and interaction of two or more autonomous rhythmic events. It was probably first described as early as the seventeenth century by the Dutch researcher Christian Huygens, who invented the pendulum clock (Pikovsky, Rosenblum, and Kurths 2001). He observed that two pendulum clocks hanging side by side from a common support synchronized their oscillations perfectly, and the pendula swung in opposite directions, due to small coupling forces transmitted through the air and/or the wall. This discovery had a strong impact on the mathematical and technological developments of that time (for a thorough review of the concept of synchronization including its historical perspective see Pikovsky, Rosenblum, and Kurths 2001); thereafter, synchronization was observed in a wide range of biological, physical, and social systems. From fireflies coordinating their flashing light patterns (Strogatz and Stewart 1993), to the sleep-wake cycle interacting with the 24 hour cycle of light and darkness (Czeisler et al. 1986), to the pendulum clocks synchronizing their oscillations, to humans clapping their hands in a crowd (Néda et al. 2000), playing sports (Silva et al. 2016), music (Keller 2013), and dancing (Ellamil et al. 2016), to cite only a few, synchronization represents a fundamental and ubiquitous phenomenon.

Synchronization can involve an external stimulus, either environmental (as in the case of the light-dark cycle, for example) or interpersonal (as in the case of music ensembles, among others). Synchronization also occurs between internal stimuli within the same individual, when two or more of a body's rhythmical oscillations interact. Examples of self-synchronization can be observed in the interaction between breathing and heart-beat (Bonsignore et al. 1995; Yasuma and Hayano 2004); in the coordination of arm and leg movements during various forms of locomotion, including walking at different speeds, creeping on arms and legs at a constant speed on a treadmill, and swimming (Meyns, Bruijn, and Duysens 2013; Wannier et al. 2001); and, in the synchronization between speech and gestures (Chu and Hagoort 2014).

Furthermore, self- and interpersonal synchronization can be observed and studied at different temporal resolutions. From the bottom up, synchronization can be analysed at the level of brain activity (Lindenberger et al. 2009; Sänger, Müller, and Lindenberger 2012; Babiloni et al. 2011; Sänger, Müller, and Lindenberger 2013) and measured in milliseconds or seconds using electroencephalographic (EEG) and magnetoencephalographic (MEG) devices, or functional magnetic resonance imaging (fMRI) techniques, respectively. At higher levels, synchronization can involve sensorimotor coordination, largely analysed through tapping tasks (for a full review see Repp 2005; Repp and Su 2013); it also involves physiological processes, as in the case of the cardiac and respiratory activity likely to be synchronized between singers during choir singing (Müller and Lindenberger 2011). At the highest levels, interpersonal synchronization can include coordination of: i) speech during conversation; ii) sound during music ensemble performances, which has been mostly analysed so far through acoustic recordings of instrumental performances (Keller 2013); and, iii) coordination of body movements. An increasing number of studies have analysed the body movements occurring between musicians during ensemble performances (for example, Badino et al. 2014; Glowinski et al. 2013; Timmers et al. 2014).

In some activities, coordination emerges spontaneously, for example when people in rocking chairs or walking together adjust their movements or stride, matching one another

(Demos et al. 2012). Sometimes, coordination of movements occurs even when people intentionally try not to coordinate their actions, if visual information between them is shared (Issartel, Marin, and Cadopi 2007). In other contexts, synchronization is intentional and pre-planned, being based on specific goals, such as during music ensemble performances. Notably, musicians in small ensembles coordinate their actions with their co-performer's actions to produce a cohesive performance; orchestral members follow the conductor's gestures; and, jazz musicians groove to the beat of a rhythm section (Keller, Novembre, and Hove 2014).

1.2 Focus and Aim of This Research

Research on interpersonal synchronization in music ensembles is still in its relative infancy. This research focuses on interpersonal synchronization between musicians during singing ensemble performances, as manifested in acoustic and electrolaryngograph recordings. Specifically, this study investigates the roles of visual contact between singers and leadership, and the evolution of synchronization and leader-follower relationships emerging across rehearsals, through three related major empirical investigations. The first investigates the effects of visual contact between singers and leadership instruction in 12 semi-professional singing duos. The second study observes the developmental aspects of synchronization and the spontaneous emergence of leader-follower relationships in a newly formed, advanced singing quintet, in relation to the musical content of the piece being rehearsed. In parallel, the second study analyses the tuning systems used by the singers and the verbal interactions between singers during rehearsal. Finally, the third study analyses whether the different patterns of synchronization resulting from altered visual contact and rehearsal are perceived by listeners with different levels of musical expertise.

These investigations address the following overarching research question:

> **Do visual contact, leader-follower relationships and rehearsal influence interpersonal synchronization between singers during singing ensemble performances?**

1.3 Hypothesis

Although this research is explorative in nature, the following overarching hypothesis has been formulated, based on previous investigations, and is considered and justified in detail in the chapters relating to each investigation:

> **Interpersonal synchronization is not affected by the presence or absence of visual contact between singers during the performance of a two-part piece with regular metre, improves with practice across rehearsals, and is associated with a mutual adaptation of leader-follower roles.**

1.4 Novel Contribution to Research

This thesis explores a number of aspects of synchronization in ensembles that have not been fully addressed in prior research, providing the following novel contributions to this field of study:

- Research on synchronization in ensembles has been mostly focused on instrumental performances, and the study of temporal coordination in ensemble singing currently lacks full investigation. This thesis centres on studies on synchronization in 12 singing duos and a newly formed advanced singing quintet.
- The role of visual cues in synchronization during ensemble performances is a topic of interest in the field of interpersonal synchronization, but findings have shown complex results (i.e., visual contact between instrumental performers might come into play as a secondary aid when the auditory feedback is restricted and the tempo irregular). This study sheds some light on the influence that visual contact might have during singing duo performances.
- It has been suggested that group roles such as leader and follower relationships might also influence interpersonal synchronization. This thesis explores how the instruction to act as leader or follower influences synchronization in duo singing, and how leader-follower relationships contextualized in terms of preceding or lagging a co-performer evolve in a singing quintet during five rehearsal sessions.
- Previous investigations have been focused on the analysis of single performance sessions, whilst the evolution of synchronization with practice remains mostly unexplored. This study analyses the developmental aspects of synchronization of a

singing quintet recorded across five rehearsal sessions from September 2017 to January 2018.

- Research so far has investigated interpersonal synchronization focusing on the coordination of timing in acoustic recordings, and/or changes in body movements, and/or physiological factors such as heartbeat, breathing, and brain activity. This project makes a novel contribution to research on interpersonal synchronization with an innovative method that investigates synchronization of note timings in parallel with the analysis of tuning to contextualise findings regarding synchronization within the wide scope of expressive ensemble behaviours.
- This dissertation represents also a novel contribution to this field of research, providing a study of the rehearsal interactions in addition to interpersonal synchronization and tuning, through a mixed design involving a quantitative analysis of interpersonal synchronization, leader-follower relationships and tuning resulting from the analysis of sound production, and a qualitative investigation of the verbal discussions between musicians during the rehearsal sessions.
- Synchronization during ensemble performances has been analysed mostly in relation to note onsets. Although coordination at note offsets might not be of primary importance to some musical instruments such as the piano (due to the use of the damper pedal), this might play an important role in ensemble singing. This study investigates interpersonal synchronization at note onsets and offsets in parallel.
- Research on the perception of interpersonal synchronization is scarce. This research contributes exceptionally to this field by observing whether the different synchronization patterns measured in the singing recordings and resulting from altered visual contact and rehearsals are reflected in the perception of synchronization by listeners with different levels of musical expertise.
- The analysis of interpersonal synchronization in ensemble singing represents a major challenge due to the difficulty of: i) separating individual voices within polyphonic recordings to evaluate the contribution of each singer, and ii) identifying tone onsets and offsets, which are not clearly distinguishable in singing but vary according to vibrato, vocal fluctuations, timbral characteristics, and onset envelopes. This research develops and tests a novel method combining acoustic analysis and electrolaryngography to detect onsets/offsets of phonation and note

beginnings/endings within a legato phrase, through an automated algorithm termed TIMEX. The effectiveness of this method for the analysis of synchronization in ensemble singing is then tested in a pilot study conducted with two singing duos. Finally, the method is successfully applied in two empirical investigations analysing the roles of visual contact, leadership and rehearsal on synchronization in ensemble singing.

1.5 Significance

Music ensembles represent one of the most tightly synchronized activities in which human beings, from amateur musicians to professionals, take part. For this reason, the adoption of a music ensemble paradigm is particularly significant to the study of interpersonal synchronization of joint actions. The features of interpersonal synchronization highlighted in this dissertation are meaningful to all human activities involving social interaction.

1.6 Thesis Layout

This thesis is structured as follows:

- **Chapter 2** presents and debates the existing literature relevant to interpersonal synchronization in music ensembles, describing its distinctive features within and between cultures, the factors underpinning and influencing the sustainment of synchronization, and the methods used for the study of synchronization. Most of this chapter has been published in D'Amario and Daffern (2017) and D'Amario, Daffern, and Bailes (2018a; 2018b; 2018c).
- **Chapter 3** presents and tests a novel algorithm, termed TIMEX, that automatically detects note onsets/offsets in ensemble singing, based on the combined application of electrolaryngography and acoustic analysis. This study is reported in D'Amario, Daffern, and Bailes (2018b).
- **Chapter 4** implements and tests the efficacy of a new protocol, based on TIMEX, to investigate interpersonal synchronization between singers during ensemble performance. The effectiveness of this protocol is evaluated in a pilot study with two singing duos, investigating the impact of visual contact and leadership instruction on interpersonal synchronization during singing duo performances. This study is reported in D'Amario, Daffern, and Bailes (2018b).

- **Chapter 5** builds on the previous pilot study, describing an empirical investigation conducted with 12 singing duos to analyse the roles of visual contact and the instruction to act as leader or follower on interpersonal synchronization. This investigation appeared in D'Amario, Daffern, and Bailes (2018c).
- **Chapter 6** describes another empirical investigation observing the evolution of synchronization and the leader-follower relationships emerging spontaneously between the co-performers across rehearsals in a newly formed, semi-professional singing quintet. The study of synchronization is contextualised within the analysis of tuning, another fundamental musical parameter that, in addition to timing, musicians vary during ensemble performances to enhance the expressiveness of ensemble performance. Thus, overall this study observes whether changes in the synchronization are supported by changes in the tuning outcomes. This study has been reported in D'Amario, Daffern, and Bailes (2018a) and in D'Amario et al. (2018).
- **Chapter 7** describes a perception study investigating whether the differences in the synchronization patterns resulting from the altered visual contact (observed in chapter 5) and rehearsals (reported in chapter 6) are perceivable by listeners with different levels of music expertise.
- **Chapter 8** summarises the main findings and the contributions of this research to the field, and presents new areas for future investigations. The chapter concludes with some final remarks on interpersonal synchronization between singers in ensemble singing.

2 Interpersonal Synchronization in Music Ensemble Performances

> *The asynchronization of simultaneous tones should be regarded as one of the vital deviations in the performance of music.*
>
> (Rasch 1988, 81)

This chapter reviews the literature relevant to interpersonal synchronization in music ensemble performance as manifested in the sound produced, to lay down the foundations upon which this thesis is built. The main characteristics of synchronization are presented between and across cultures alongside the cognitive, motor and physiological factors underpinning synchronization. The effects of certain high-level factors that may affect the performance of synchronization, and the fundamental musical parameters important to contextualise synchronisation, are reviewed. The perceptual aspects of synchronization are then presented. This chapter concludes with a review of the methods and techniques used for the measurement and assessment of synchronization during ensemble performances.

2.1 Introduction

2.1.1 Synchronization in Western Classical Music

Timing within a music ensemble performance varies within and between players, establishing asynchronies between members of an ensemble. Part of this variability in Western music is intentional and pre-planned, relating to the musical score or shared intentional deviations from the score in support of expressive goals, such as deliberately slowing the tempo at the end of the piece (Phillips-Silver and Keller 2012) or delaying some notes as a means of emphasis. A certain amount of this variability is also unintentional, due to technical and/or expressive complexity and noise during the cognitive-motor processes (Ragert, Schroeder, and Keller 2013). Musicians generally try to limit and control the extent of these inter-performer temporal fluctuations through individual practice and collaborative group rehearsals, with the purpose of establishing shared performance goals based on knowledge of the musical structure and the playing style and expressive intentions of the co-performer(s) (Williamon and Davidson 2002; Ginsborg, Chaffin, and Nicholson 2006).

The variability in note onset asynchronies between performers in Western Classical professional ensembles is typically very small, in the order of tens of milliseconds, and decreases with increasing tempo (Rasch 1979; Rasch 1988). The average value of absolute asynchronies and the standard deviation of signed asynchronies are both typically between 30-50 ms during expressively performances; mean signed asynchrony is typically close to 0, falling often between -5 to +5 ms (Keller 2014; Goebl and Palmer 2009; Timmers et al. 2014; Bishop and Goebl 2015; Zamm, Pfordresher, and Palmer 2015). These slightly different values reported might be understood in light of the different types and sizes of the ensembles, the different music pieces used for the studies, and/or the different methods of data collection. Nevertheless, these levels of coordination are maintained through iterative temporal adjustments: people may adapt the timing of their finger tapping to that of an autonomous timing source such as a metronome in tapping tasks (Repp and Su 2013); musicians may correct the tempo to one of the co-performers or each player may adjust the tempo for the temporal fluctuations of the other (Goebl and Palmer 2009).

Interpersonal coordination of timing has been also investigated among jazz players. Jazz performances are characterized by meaningful variations of tight and loose coordination between the pairs of musicians, which suggest dynamic interpersonal relationships (Doffman 2008; Doffman 2013; Doffman 2009). These timing fluctuations between jazz musicians may provide the music performance with its vital drive, named “groove”. The term commonly refers to the quality of music that urges the listeners to move (Madison 2006; Janata, Tomic, and Haberman 2012). More recently, the tendency to lead/lag, among other aspects, was investigated in six jazz trio combinations (of one saxophonist, two bassists and three drummers during the performances of three popular songs) by calculating the signed asynchronies between the drummer’s ride cymbal taps and the bassist’s onsets (Hofmann, Wesolowski, and Goebl 2017). Results showed a mean signed asynchrony of 2.1 ms computed across the six jazz trio combinations, which is in line with that found in Western Classical ensembles; moreover, researchers found that this synchrony was not dependent on the song played, but on the combination of the musicians’ playing styles.

Furthermore, the analysis of interpersonal synchronization has been mostly conducted investigating small ensembles such as duos (Goebl and Palmer 2009; Palmer et al. 2013; Keller and Appel 2010; Zamm, Pfordresher, and Palmer 2015; Bishop and Goebl 2017; Bishop and Goebl 2015) and quartets (Timmers, Endo, and Wing 2013; Timmers et al. 2014; Badino et al. 2014; Glowinski et al. 2012). Synchronization characteristics might change with an increasing number of musicians performing together. Investigations might consider larger ensembles, such as quintets, which are still common in the Western Classical tradition. Analysis remains feasible in a group of this size, but different patterns of relationships between musicians might be evident given the much more complex context.

Moreover, very few studies have analysed temporal coordination within singing duos, reporting a mean signed asynchrony value of less than -5 ms when performing the same melody in unison (Palmer et al. 2013). Analysis of interpersonal synchronization in ensemble singing is currently lacking investigation, probably due to the complexities of the analysis (for a full review, see sections *2.5.1, 2.5.2* and *2.5.3*). Nevertheless, interpersonal synchronization in singing ensembles is distinct from any other instrument due to the inclusion of lyrics. Singers usually communicate a text in addition to a musical expression. During singing ensemble performances, singers have to ensure synchronization within the context of a specific text, which is also heavily dependent on the acoustic conditions of the room (Brereton 2014). For example, the placement of plosive consonants in relation to the onset of phonation might change between a small room or a large cathedral adding to the complexities of interpersonal coordination of timing. Singers might focus on the synchronization of a plosive consonant in small rooms, but they should consider the delay of the sound before the onset of phonation in the context of a large cathedral. Due to the extensive possibilities available to the human voice and the complexities of multiple voices performing together, the analysis of temporal coordination in ensemble singing has the potential to deepen our understanding of interpersonal synchronization.

Finally, research has been mostly focused on synchronization at onsets (Goebl and Palmer 2009; Loehr and Palmer 2009; Timmers et al. 2014; Zamm, Pfordresher, and Palmer 2015; Hove, Keller, and Krumhansl 2007; Bishop and Goebl 2015; Bishop and Goebl 2017; Hofmann, Wesolowski, and Goebl 2017); little is known about coordination at offsets. This may not be a meaningful measure for pianists given their use of the damper pedal, but might be relevant to other types of music ensembles, including singing ensembles, where musicians do try to synchronize offsets as well as onsets.

2.1.2 Ethnomusicological Perspectives

In addition to the study of Western music described above mainly from the perspective of Western Classical music (although jazz music was also commented on), a few ethnomusicological investigations have focused on the interpersonal coordination of timing in music ensembles within and between cultures. The ethnomusicological approach to synchronization, though still in its infancy, has the potential to reveal the cultural aspects of interpersonal coordination identifying cross-cultural similarities and differences. Martyn Clayton, a pioneer of ethnomusicological studies in interpersonal coordination, conducted several investigations within Indian and Afro-Brazilian cultures, and observed a number of cases of music "entrainment". This phenomenon refers to two independent rhythmical systems that synchronize, interacting with each other to such an extent that they eventually assume the same or related period; in other words, two independent rhythms that stabilise and reassert this stabilization if that synchrony is momentarily stopped (Clayton 2012; Clayton, Sager, and Will 2005). This means that synchronization in itself is necessary but not sufficient for a relationship to be described as entrainment: only when synchronization is re-formed after disruptions, then two oscillations are entrained.

Cases of music entrainment were observed between independent groups of musicians during Afro-Brazilian Congado performances, a form of ritual processional music (Lucas, Clayton, and Leante 2011). Research also showed that the close spatial proximity of the groups was a necessary but not sufficient condition for entrainment. In one case, in fact, the two groups under investigation were in close proximity but managed to avoid entrainment when visual contact was absent between the groups; this result suggests that visual contact might also have an important role to entrain in music during Congado rituals among Afro-Brazilian communities.

A study investigating Indian Classical music revealed complex relationships between tanpura players in an Indian classical ensemble (Clayton 2007b). Whilst the musicians' intentions were to keep independent tanpura rhythms, observational analysis of the performers' hand gestures revealed cases of unintentionally tight coordination. When the accompanist fixed their visual attention on the soloist's back or shoulder, the tanpura rhythms stabilised in 3:2 relationships (the period of one plucking pattern was roughly 3 seconds, whilst the other was 2 seconds), suggesting the importance of visual contact between musicians in Indian classical ensembles.

More recently, interpersonal synchronization between musicians was investigated in the context of Transylvanian (the central region of the current Romania) village music. Research found *aksak* patterns (rhythmic sequences of short and long beats executed based on uninterrupted reiterations) and entrained interactions between two professional Gypsy players performing a repertoire known as “Gypsy song of sorrow” mostly performed during local celebration events (Clayton 2015; Bonini-Baraldi, Bigand, and Pozzo 2015). These studies also measured asynchronies ranging from a few milliseconds to almost a second, which the authors described as small asynchronies, less than 10 % of the beat duration, and large asynchronies, from 20 % to 50 % of the beat duration, depending on the musical structure of the melody in relation to the accompaniment.

Notably, these studies exploring Transylvanian music also observed a tendency for the violinist to play ahead of the viola. This finding is important in light of the results observed in Western Classical string quartet performances by Timmers, Endo, and Wing (2013) and Timmers et al. (2014), showing a more complex division of roles among co-performers rather than the clear leadership of the violinist during the repeated performances of a short excerpt of the first movement of the string quartet op. 77 n. 1 by Joseph Haydn (see section *2.2.2* for more detail). Such apparently contrasting results can be understood on the basis of the different social role played by the violinist, or the music pieces performed. As the authors argue, the violinist in the Transylvanian music is the ensemble leader, responsible for, among other aspects, managing the deal with the clients, choosing the tunes, and setting the tempo. This role of the violinist, specific to the Transylvanian music and absent from the Western Classical ensemble culture, might have induced the leadership during the performance. The different structural characteristics of the pieces performed might have also elicited different results between the studies: the melody less ambiguously assigned to the violin in Transylvanian music, compared with the Haydn’s piece, might have induced the violinist to lead.

In summary, this section has reviewed recent ethnomusicological studies, demonstrating the importance of cross-cultural perspectives on the study of interpersonal synchronization across cultures. The next section presents the cognitive-motor and physiological mechanisms that contribute to the fulfilment of synchronization in ensemble performance.

2.1.3 Cognitive-Motor Processes and Physiological Mechanisms

A theoretical framework has been proposed including three cognitive-motor processes that enable a tight interpersonal coordination in ensemble performances, named: i) *anticipatory auditory imagery,* ii) *prioritized integrative attention,* and iii) *adapting to others' action timing* (Keller 2008; Keller 2013). The first process, *anticipatory auditory imagery,* refers to the ability to anticipate one's own sound and the co-performers' sound, creating auditory and motor imagery of their auditory effects; this auditory imagery ability depends on the level of musical experience (Keller and Koch 2008). The second process, *prioritized integrative attention*, refers to an attentional strategy that musicians employ to facilitate cohesive and precise performance. Musicians pay attention to their own actions and to the co-performers' actions, whilst assessing the overall incoming sound from the ensemble. The third process, *adapting to others' action timing*, is an important ability in ensemble performance, and refers to the constant temporal adjustment of one's own timing with the co-performer(s)' timing. Temporal adjustments are necessary to master intentional expressive tempo changes or unintentional temporal fluctuations due to noise in the cognitive motor processes.

These continuous temporal adaptations are sustained by two independent error correction processes, named phase correction and period correction (Keller 2008). Phase correction refers to the automatic process that adjusts the way a sequence of pulses, such as quarter beats, produced by a timekeeper in one performer is aligned with a sequence of pulses in a co-performer; thus, greatly supports precision in interpersonal synchronization (Keller 2014). Period correction, by contrast, refers to the controlled adjustments of the duration of each timekeeper interval (i.e., the duration of each pulse) on the basis of previous information; it is not automatic but requires conscious and explicit attention, control and awareness by the musicians (Repp and Keller 2004; Repp 2001). Period correction is mostly needed with obvious tempo change, whilst phase correction is continuously triggered to adapt to small temporal fluctuations.

Findings regarding error correction processes are mostly based on tapping tasks, in which participants tap along to rhythmic sequences of auditory stimuli (for a review, see Repp and

Su 2013; Repp 2005). Recently, they have also been investigated in two professional string quartets by Wing, Endo, Bradbury, et al. (2014). In one quartet, the first violinist showed fewer adjustments to the co-performers than the others' adjustment to her. In another quartet, researchers found no difference in the adjustment patterns between the first violinist and the other members of the ensemble. These findings suggest different strategies used during the performance, i.e. led by the first violin *versus* a more democratic approach.

In addition to the studies investigating the cognitive processes that support synchronization, an increasing body of research has explored the physiological factors that can constrain or affect interpersonal synchronization during ensemble performances, from brain activity to body movements' coordination. Studies in this field suggest that a network of motor brain areas, tightly synchronized and interacting within and between brains, allows the execution of an efficient social coordination during ensemble playing (for a review of the neurophysiological mechanisms allowing real-time interpersonal coordination see Keller, Novembre, and Hove 2014). It has been shown that interpersonally synchronized activities such as duo guitar performances were preceded and accompanied by between-brain oscillatory activity (Lindenberger et al. 2009; Sänger, Müller, and Lindenberger 2012), with coupled brain activity in the alpha and beta frequency ranges between the guitarists (Sänger, Müller, and Lindenberger 2013). The directionality of such between-brain couplings discriminated musical roles such as leader and follower (Sänger, Müller, and Lindenberger 2013). Modulations in the cortical activity have also been found during saxophone quartet performances, with alpha power density values decreasing in amplitude in several cortical regions, whereas power density values enhanced within narrow high-frequency bands during ensemble playing compared to a resting state (Babiloni et al. 2011). These results suggest a relationship between temporal coordination of actions during ensemble performances and cortical activity. Overall, little is still know about the specific neurophysiological mechanisms underpinning interpersonal synchronization during ensemble performances (Babiloni et al. 2011). This is probably due to the limitations of the technology involved; for example, electroencephalogram (EEG) and magnetoencephalogram (MEG) offer high temporal resolution (ms), but low spatial resolution (cm), whilst fMRI offers low temporal resolution (s), but high spatial resolution (mm) (Babiloni et al. 2011). In addition, metal objects are not permitted in rooms using MEG and fMRI, and head-movements should be limited during EEG to avoid artefacts in the recordings. The limitations of the current methods used to track neurophysiological data during ensemble performances inevitably limit the research in this field.

At a higher temporal level, interpersonal synchronization can involve the temporal coordination of other physiological elements, such as breathing and cardiac activity. Investigations focused on the musical domain have shown a tight synchronization of such elements during ensemble performances. For example, patterns of cardiac and respiratory synchronization have been found in choir singing, with phase synchronization between respiration and heart rate variability increasing during singing compared with a rest condition, and also increasing when singing in unison rather than singing with multiple voice parts (Müller and Lindenberger 2011). Tight coordination of cardiac activity between singers has also been reported in a study investigating heart rate variability of singers in relation to the musical structure of the piece being sung. Results show a connection between song structure, heart-rate and respiration, and synchronization in frequencies and phases of respiration and cardiac cycles between singers (Vickhoff et al. 2013).

Furthermore, the analysis of body movements is particularly relevant to the analysis of interpersonal synchronization, since musicians' movements play a crucial role in the exchange of information during ensemble performances. Musicians, in fact, commonly make use of body gestures, such as head nods, to communicate piece onsets. This has led to a rising corpus of research investigating cueing gestures, such as head nods and body sway, and defined ancillary movements (Nusseck and Wanderley 2009). Such movements are not strictly involved in the sound production, but support the musical performance by facilitating interpersonal communication, for example communicating elements of the formal structure of the piece (Davidson 2001; Wanderley et al. 2005; Williamon and Davidson 2002). It has been shown that musicians' movements are more predictable when they play in an ensemble than alone (Glowinski et al. 2013). Musicians' movements during a performance differ with varying degrees of shared musical intentions between musicians (Glowinski et al. 2015). Synchronization of movements has also been investigated in relation to the level of expertise of the musicians during a cross-cultural choir workshop involving four South-African experts and four Finnish novices (Himberg and Thompson 2009). Results show that the experts' movements were more synchronized than the novice group.

Certain spatio-temporal characteristics of cueing gestures enable synchronization with visual gestures; for example, in piano-piano duos, violin-violin duos, and violin-piano duos, peak acceleration in a leader's head-nodding gestures represents beat position, whilst gesture duration and periodicity of head and bowing hand gestures represent tempo (Bishop and Goebl 2017). The finding regarding points of peak acceleration in head-nods to cue beat position has been corroborated by Bishop and Goebl (2018), showing also that interpersonal synchronization at onsets improved as smoothness and magnitude of the gestures increased, and the prototypicality (i.e., the similarity between the gestures produced by co-performers, evaluated by assessing how similar each gesture was to all other gestures within the data set) decreased. Furthermore, research has shown that pianists' head movements in piano duos became more synchronized when auditory feedback was reduced, and that the fingers of pianists who were acting as leaders were raised higher than those of the followers (Goebl and Palmer 2009). The bow movements of the first violin in a professional string quartet was representative of the primary leadership of the violinist, who set the tempo of the piece through anticipatory movements of the bow (Timmers et al. 2014). The bow speed was faster when preparing for longer notes, which the authors speculate might be related to the musician's intentions to emphasize particular notes through intensity and duration. More recently, movements have been investigated in relation to jazz duos, comprising a range of instruments, i.e. saxophone, piano, double bass, electric bass, drums, trumpet, guitar, flute, clarinet, violin and cello (Eerola et al. 2018). Results suggest that ancillary body motion may vary according to the temporal regularity of the music, and may be more important for co-performer communication in non-pulsed, free improvisations, than in standard jazz with a regular pulse.

In summary, interpersonal synchronization is supported by a complex and highly interconnected range of cognitive, motor, and physiological mechanisms that support synchronization in ensembles from the lowest levels of the temporal hierarchy, including brain activity, to the highest levels, comprising the body movements. The simultaneous investigation of such different aspects has the potential to provide a deep understanding of interpersonal coordination, revealing the cognitive and (neuro)physiological processes supporting interpersonal synchronization in music ensembles.

2.2 Factors Affecting Synchronization

In addition to the cognitive, motor, and physiological mechanisms described above, research has revealed a number of external factors that may affect the production of interpersonal synchronization, such as the presence/absence of visual contact, leader-follower relationships, and rehearsal. The roles of visual contact, leader-follower relationships, and rehearsal in interpersonal synchronization between musicians in ensemble performances are reviewed in detail in the following sections.

2.2.1 The Role of Visual Contact

Visual contact between members of an ensemble is a key element of ensemble performance that might affect synchronization in joint music performance, and a number of studies have investigated the nature of these effects. Investigations of unintentional interpersonal communication in non-musical contexts have demonstrated an effect of visual contact on interpersonal entrainment (Oullier et al. 2008). Studies analysing the role of visual contact in musical scenarios have demonstrated that eye contact is often used in popular music bands (Kurosawa and Davidson 2005) and piano duo collaborations (Blank and Davidson 2007). A study investigating visual contact between members of a band and a conductor reported that ensemble musicians during a performance looked at a videotaped conductor for 28 % of a performance duration (Fredrickson 1994). It has also been found that the frequency of visual contact among string quartet players did not change in relation to the stress associated with the performance setting, i.e. rehearsal setting *versus* public recital (Biasutti et al. 2016). But the frequency of visual contact increased across different stages of rehearsals in piano duos, as pianists became more familiar with the music and felt more free to raise their eyes from the score (Williamon and Davidson 2002). A number of studies have revealed that visual cues improve the communication of interpersonal intentions between musicians (Dahl and Friberg 2007; Castellano et al. 2008). Qualitative investigations have suggested that eye contact also improves synchronization in musical ensembles (Clayton 2007a; Williamon and Davidson 2002).

A few quantitative studies analysing the benefits of visual contact for interpersonal synchronization in the music ensemble context have elicited complex results. Keller and Appel (2010) found that the presence or removal of eye contact did not markedly affect synchronization between pianists in duo performances, as indexed by the median of signed and unsigned asynchronies, calculated by subtracting the onset times of the *primo* part from the *secondo* part. However, higher variability of temporal coordination, as indexed by the coefficient of variation (CV) of signed asynchronies, was found in the presence of visual contact compared with when visual contact was removed, which the researchers speculated could be because the musicians may have focused more on expressive timing variation. Research also suggests that visual cues between pianists are more important when auditory feedback is limited compared with full auditory feedback (Goebl and Palmer 2009), and that eye contact might be important for the temporal coordination between pianists (Kawase 2014). The different results reported by Keller and Appel (2010), Goebl and Palmer (2009), and Kawase (2014) might be explained in relation to the characteristics of the musical stimuli being performed, as discussed by Bishop and Goebl (2015). In the first two studies, the authors made use of pieces with a regular metre, while the latter utilized a rhythmically complex piece featuring tempo changes and long pauses. The effect of visual contact in relation to tempo change was tested by Bishop and Goebl (2015), demonstrating that eye cues positively affect temporal synchronization during piano-piano duo and piano-violin duo performances when following long pauses in the music. These results suggest that visual contact between pianists or piano-violin players might come into play as a secondary support in improving synchronization when auditory feedback is limited or musical timing is irregular.

The role of visual contact in singing ensembles currently lacks investigation. This aspect might be particularly relevant to singing as a singer's body is their instrument. Whilst some instrumentalists such as piano players might have the additional factor of looking at their musical instrument, which is physically independent from their own body, a singers' vision might be mostly focused on the score and/or co-performer(s)' and conductor's body gestures during singing performances.

2.2.2 Leader-Follower Relationships

Synchronization in joint music performance may also be influenced by group roles such as leader-follower relationships between members of a musical ensemble. Research in this context has investigated both leadership emerging spontaneously during ensemble performances, and the effect of the instruction to act as leader or follower on interpersonal synchronization.

A number of case studies have recently investigated leadership emerging spontaneously during quartet performances by analysing body movements (e.g., head and instrument's bow) in relation to acoustic cues. As briefly introduced earlier in section *2.1.2*, studies conducted by Timmers, Endo, and Wing (2013) and Timmers et al. (2014) have shown a complex pattern of leader-follower roles between musicians during string quartet performances, rather than a traditional role division of leadership characterized by the artistic attribution of leader to the first violin. Analysis of the audio onsets, based on the cross-correlations between interbeat intervals at different lags, demonstrates mutual adaptation of roles between the first violin and both the viola and the cello (Timmers, Endo, and Wing 2013); bidirectional dependence between the second violin and cello, and the second violin and viola (Timmers et al. 2014); and, a unidirectional dependence of viola on the first violin, and of the first violin on cello (Timmers et al. 2014). On the other hand, the leadership of the first violin was revealed in the analysis of the bow movements preceding the performance of the first onset (Timmers et al. 2014). A study conducted by Glowinski et al. (2012) demonstrated also the relative leadership of the first violin, investigated through the analysis of the movements of the musicians' heads towards a common point of reference. Results show that the first violin exhibited the highest number of driving forces, an indicator of the relative importance of the musician, although that of the other musicians remained close to the first violin. A study conducted by Badino et al. (2014) tried to force the unidirectional communication between the first violin of a string quartet and the co-performers by applying temporal and dynamic changes to the score, known only to the first violin, across repeated performances. Results show that when perturbations were introduced, unidirectional influence from the leader decreased, suggesting that leadership might depend on the sharing of knowledge between performers.

The study of leader-follower relationships in the context of ensemble singing is also of great interest. A recent qualitative study based on individual, semi-structured interviews and focused on a professional singing octet found that leadership was democratically spread across the ensemble, and reported that each singer felt the ensemble was not led by a single singer (Lim 2014). Similarly, a qualitative investigation analysing an 11-men singing ensemble showed that all members of the ensemble felt empowered to take a lead (Page-Shipp, Joseph, and van Niekerk 2018). However, leader-follower relationships, as quantified by the tendency to precede/lag a co-performer at onsets/offsets, have not been fully understood in the context of ensemble singing. Further investigations might shed more light in this respect.

As anticipated above, leader-follower relationships have also been investigated by assigning specific group roles. Goebl and Palmer (2009) analysed leadership instruction in piano duos performing three pieces with a different note ratio between the parts (i.e., 1:1, 2:1, and 1:2). They found that when auditory feedback of the pianists was limited, pianists performing more notes preceded the co-performers playing fewer notes, irrespective of the leadership instruction. Zamm, Pfordresher, and Palmer (2015) further analysed synchronization in piano duets, showing a compensatory timing behaviour between pairs of pianists performing the same melodies in a round, characterized by a delay in temporal attack between one pianist who begins and is assigned the role of leader, and a second pianist who enters later and is assigned the role of follower. The study reports, in fact, that the followers' onsets precede those of the leader, showing a directionality that is opposite to the researcher's instructions and to the musical structure. Although the analysis was not able to identify whether this directionality was due to the follower striving to catch up, or to the leader lagging behind, a compensatory behaviour is evident. Furthermore, researchers' requirements to keep a fixed tempo by listening to a metronome before each trial, and to instruct the designated leader to be responsible for determining the tempo, might have affected these leader-follower relationships.

Although leader-follower roles are generally conceptualized as social roles, rather than in terms of performance timings, the above findings overall suggest that investigating the anticipation-delay of onsets by performers within an ensemble is a valuable indicator of group roles. The studies conducted so far to understand music roles through the analysis of synchronization between musicians during ensemble performances have also highlighted the complexity of the phenomenon and the need for future investigations. For example, the effect of the instruction to act as leader or follower without a focus on time-keeping or leadership clearly induced by the score is not fully understood. Investigation to this end would be particularly beneficial for singing ensembles, since the literature analysing temporal coordination has been mostly focused on instrumental ensembles.

2.2.3 Evolution across Rehearsals

Whilst there is a strong body of research into interpersonal synchronization which analyses single performance sessions, the developmental aspects of synchronization in ensemble singing remain mostly unexplored. Members of professional ensembles synchronize their entrances with near-perfect precision with the other co-performer(s), but this may require practice over several rehearsals to be achieved (Williamon and Davidson 2002). Therefore, investigation of the evolution of temporal coordination between musicians across rehearsals is of interest to music pedagogy, aimed at refining rehearsal strategies, and music psychology and psychology research, in terms of understanding social interaction.

Observational studies of small ensembles have demonstrated ways in which preparation for performance requires musical and social coordination, generally achieved through a framework of rehearsals and performance goals, with variation between groups of different type, size and familiarity (Blank and Davidson 2007; Davidson and King 2004; Goodman 2002; Williamon and Davidson 2002). As part of a study of ensemble rehearsal approaches, Chaffin and Imreh (2002) categorised rehearsal tasks as ‘basic’, interpretive’, ‘expressive’ and ‘strategic’. This framework was later adapted and applied in studies of ensemble rehearsals, including that of Ginsborg, Chaffin, and Nicholson (2006), a longitudinal study of rehearsal of a professional voice and piano duo. Using verbal utterances to track the focus of the rehearsals, they characterised work on tempo, note and rest durations, pitch and intonation as ‘basic’ musical dimensions. Over the course of the study they observed a shift from these more 'basic' tasks in early rehearsals to a greater emphasis on ‘interpretive’ tasks (such as expressive intentions) in later sessions. This framework was also used to explore differences in rehearsal approaches in a small-scale study (four duos) of newly formed and established student and professional ensembles (Ginsborg and King 2012). There were no differences found in verbal utterances referring to 'basic' musical dimensions relating to expertise or familiarity. In light of these findings, future research into the evolution of the synchronization patterns during rehearsals should be conducted in combination with the analysis of the verbal interactions during rehearsals, in order to investigate the rehearsal strategies adopted to support interpersonal synchronization between musicians.

In addition, a study about the perception of asynchrony conducted by Mossbridge et al. (2006) showed that a single exposure to perceptual tasks (i.e., detecting asynchrony and discriminating the temporal order of two onsets not perfectly synchronized) increases the ability to detect asynchrony. This suggests that rehearsals that focus specifically on synchronization might increase the interpersonal synchronization between musicians during ensemble performances. However, the evolution of synchronization across rehearsals in music ensembles has not yet been fully investigated.

2.2.4 Other Factors

In addition to the factors described in the previous sections, research has revealed a number of other elements that may also affect interpersonal synchronization. The impact of the rhythmical complexity of the piece being performed, for example, has been also the focus of an increasing number of studies (Loehr and Palmer 2009; Goebl and Palmer 2009; Zamm, Pfordresher, and Palmer 2015). A study conducted by Loehr and Palmer (2009) has analysed the modality of sensory information occurring between beats by manipulating the number of notes within each crotchet beat. Results show that pianists performing melodies with a metronome were less synchronized when playing melodies in which the crotchet beats were subdivided by adding quaver notes, compared with when the crotchet beats of the melodies were not subdivided. In addition, pianists performing duets with different note ratios between the two musical parts (1:1, 1:2, and 2:1 ratio) were best synchronized when the lower part was playing fewer notes (2:1 ratio) and less synchronized when playing more notes (1:2 ratio), under conditions whereby the upper parts heard only themselves whilst the lower parts heard both parts. Furthermore, in a study focused on piano duos (Zamm, Pfordresher, and Palmer 2015), research has also investigated synchronization in relation to specific structural characteristics of the piece, i.e. the effects of performing the same melody in unison or in a round, which is characterized by a delay of the entry of the second performer. Findings demonstrate that pianists were more synchronized when playing the same melody in unison than in a round, suggesting that the non-simultaneous entrance might decrease the degree of synchronization. The effect of the rhythmical complexity of the piece has not been fully analysed in larger ensembles. Further investigations might shed some light on the effects of the rhythmical complexity of the piece on synchronization between musicians in larger ensembles.

In addition to the role of the piece, the impact of other factors has been investigated. Synchronisation might be affected, for example, by the auditory feedback from co-performer(s), the familiarity with co-performers' playing styles, and the musicians' levels of ensemble expertise (Goebl and Palmer 2009; Keller and Appel 2010; Keller, Knoblich, and Repp 2007; Loehr, Large, and Palmer 2011). The study of the effects of the such elements is beyond the scope of this thesis, therefore the roles of these variables are not reviewed in detail here.

Having reviewed the factors affecting synchronization, the thesis goes on to present the context of interpersonal synchronization within expressive performances, describing deliberate variations of other fundamental musical parameters that, alongside interpersonal timing fluctuations, are introduced to attain mastery in ensemble performance.

2.3 Synchronization in the Context of Expressive Performances

As noted by Rasch (1988, 81), the interpersonal asynchronization of simultaneous tones in music ensembles represents a fundamental deviation from the perfect vertical alignment of tones prescribed exactly in the score, because it contributes to the expressiveness of music performance. In addition to timing deviations, ensemble musicians introduce variations of other fundamental musical parameters (i.e., intensity, articulation, timbre, and tuning), with the aim to communicate information about musical structure of the piece and expressive intentions to other members of the ensembles and to the audience (Keller 2014).

To achieve optimal blend, musicians in ensembles vary the sound intensity. The intensity profiles of a solo performance may change when performing in an ensemble as a consequence of the interaction with other performers, as demonstrated in a study analysing the pianist's dynamic fluctuations in the same solo and ensemble performance of Chopin's Cello Sonata op. 65 (Goodman 2002). The intensity fluctuations result from a range of factors typical of the ensemble context. For example, the so-called "self-to-other ratio" (i.e., the extent to which singers hear their sound amongst the co-performers' sound) may induce variations in intensity (Ternström 2003). The need to blend the idiosyncratic characteristics of the individual voice with those of the co-performers' voices also may cause intensity deviations.

Similarly, musicians in ensembles may change the timbre (i.e., the spectral envelope of the sound) to enhance or break ensemble cohesion. Singers, for example, may change the amplitude of formants [i.e., a peak amplitude region in the singer's frequency spectrum

(Sundberg 1987)] to blend to varying degrees with the co-performers during choral performances (Ternström 2003). Coordination of articulation (i.e., the type of transition between notes) is also required in ensembles; musicians may be required to match or mismatch legato or staccato for expressive reasons.

Finally, another fundamental musical parameter that, in addition to interpersonal synchronization, contributes to the expressiveness of ensemble performance is tuning, representing an essential characteristic of good choral singing practice, at the forefront of critical reviews, directors' manuals and singing tutors (Hansen 1964)**.** Beyond the importance of pitch matching, whereby singers produce accurate unison singing within their respective parts, in *a cappella* part singing there is the additional issue of which tuning systems and temperaments should and are employed for a group to be 'in tune' and cohesive. The choice of the musical temperament is challenging, since the tuning of a specific pitch often depends on the harmonic context in which it occurs (Ternström 2003). Exploring predicted pitch drift in three specially composed pieces, Howard (2007b) found that when key modulation occurred even over a very short piece, in a single performance by one quartet, the singers had a tendency to drift in pitch. He also found that the singers of a different quartet drifted beyond the just intonation prediction (based on acoustically pure intervals, resulting from the overtone series) and a long way from equal temperament (in which the octave is divided into 12 semitones of equal size) in two performances of one of the pieces used from the prior study, called "Exercise 3" (Howard 2007a). On the other hand, investigating tuning practices in *a cappella* part singing, Devaney, Mandel, and Fujinaga (2012) found no evidence of pitch drift in an exercise written by Benedetti in the sixteenth century to illustrate potential pitch drift associated with just intonation, when performed by four expert three-part ensembles. They hypothesised that this was due to the shortness of the exercise and the likelihood of retaining a pitch memory for the start of the piece throughout the eight-bar excerpt. More recently, multitrack recordings of a choir with 23 singers were collected to analyse the impact of altered virtual room acoustic conditions (including different reverberation times, and a dry condition) on intonation, tempo, and timing precision in choir singing. Results demonstrate that intonation was not markedly affected by simulated room acoustics, whilst tempo was slower and timing precision decreased in the condition where participants sang in virtual rooms of relatively large sizes with long reverberation times (Fischinger, Frieler, and Louhivuori 2015).

To summarize, synchronization is a fundamental expressive parameter that, alongside intensity, timbre, articulation and tuning, musicians deliberately co-vary to attain mastery in ensemble performance. The significance of these inter-individual co-variations suggests the importance of contextualising interpersonal synchronization within the characteristics of the other parameters, for example by evaluating results of the analysis of synchronization in light of the analysis of tuning, which is of particular importance to *a cappella* singing ensembles.

Having described elements related to the performance of synchronization, it is now of interest to understand how it is perceived. The next section reviews the literature regarding the perception of synchronization, highlighting factors that may influence the perception of asynchronies.

2.4 Perception of Synchronization

Empirical investigations have been conducted to analyse the perceptibility threshold of auditory asynchronies and the factors that might affect asynchrony perception. Wallach and his colleagues found that two identical sounds, of which one was delayed, were perceived as a single fused sound when presented in close succession (Wallach, Newman, and Rosenzweig 1949). This phenomenon, defined as the *precedence effect*, was found to vary by the type of sound; fusion occurred when the delay between the two sounds was up to 5 ms for short clicks, and up to 40 ms in the case of complex sounds such as speech and music (Wallach, Newman, and Rosenzweig 1949). Research also shows that pure tones lasting 5 ms and temporally offset by 30 ms or more tend to be perceived as separate auditory streams (Bregman and Pinker 1978). Mean stream segregation ratings, which quantify the extent to which two tones are perceived within the same stream, are correlated with the tones' frequency distance: the further apart their frequencies, the less likely they are to be heard in the same stream (Bregman and Pinker 1978). Listeners are more sensitive to detecting onset rather than offset asynchronies of multicomponent sounds, and perceive onset asynchrony more easily for complex sounds with components spaced harmonically rather than at equal intervals in logarithmic frequency, given the same number of components (Zera and Green 1993).

A few studies have investigated listeners' discrimination of the temporal order of sonic events, analysing their ability to identify the correct order of two onsets, i.e. which onset came first. The threshold of asynchrony to detect the temporal order of two asynchronous tones for highly trained listeners is between 15 and 20 ms, and this is mostly independent of the frequency (high or low) and bandwidth (narrow or wide) of the sounds (Hirsh 1959). The temporal order discrimination threshold was further investigated by Pastore, Harris, and Kaplan (1982), who demonstrated that it is strictly related to the stimulus duration, and can be as small as 5 ms if the duration of the sinusoidal stimuli is as small as 10 ms. Recently, the hypothesis has been tested that the threshold for the detection of asynchrony between complex music sounds such as piano tones will be higher than for steady-state sounds in case of highly trained musicians (Goebl and Parncutt 2001; Goebl and Parncutt 2003). Researchers found that this discrimination depends on the tone type, improving with more artificial tones compared with acoustic piano tones. The threshold was around 20 ms for pure tones and around 30 ms for acoustic piano tones spanning an interval of an octave or a major seventh, regardless of whether the pair began with the higher or lower tone. Research also found that this temporal order discrimination threshold was related to the music expertise of the listeners, reporting that nine musically untrained listeners (with zero to seven years of playing an instrument) were not able to discriminate the temporal order (Goebl 2003). The research above identified the threshold of temporal order discrimination in relation to pure tones and acoustic piano tones; discrimination in a more realistic scenario such as music ensembles is still to be investigated.

Research has also investigated listeners' sensitivity to asynchrony in relation to the relative intensity of the sounds. The perception of major sixth dyads was tested by manipulating simultaneously the relative timing and intensity of two tones by ±54 ms and ±20 MIDI units (Goebl and Parncutt 2002; Goebl 2003). Researchers found that the detection of asynchrony was less reliable when the louder upper tone entered before rather than after a softer, lower tone. They argued that this might be due to the effect of a forward masking, where a louder-anticipating sound masks a softer-following sound. In a following experiment, researchers investigated the influence of asynchrony and intensity in the perceptual salience of individual voices in multi-voiced musical context, comprising three-piano tone chords, sequences of piano chords, and the first nine bars of Chopin's *Ballade* op. 38 (Goebl and

Parncutt 2003; Goebl 2003). Results show that intensity was the dominant cue, whilst the effect of note asynchrony on the perceived dominance of a particular tone was comparatively marginal.

In addition to the study of the perception of individual tones, research has recently investigated the perception of synchrony in the context of string quartet and jazz trio performances. Perception of between-player asynchrony variability, for example, has been analysed using computer-simulated string quartet performances of a short excerpt from Haydn op. 74 no. 1 (Wing, Endo, Yates, et al. 2014). Within-player timing variability and correction gain (i.e., the size of the adjustments relative to the asynchrony) were manipulated in two separate experiments. Results showed that listeners without a specialized music training were sensitive to the variability of note onset asynchrony and degree of correction gain, when judging the level of togetherness in quartet performances. This study identified the need for future research to consider variation in dynamics and more complex rhythms, since this might affect the perception of asynchrony variability. In addition to the study of the perception of synchronization in string quartets, a recent study has analysed the perception of synchronization of a pool of jazz trio recordings, comprising the original performance featuring asynchrony up to 26 ms, and also recordings manipulated with increased and decreased asynchronies (Hofmann, Wesolowski, and Goebl 2017). Results show that listeners, irrespective of their music training, preferred recordings with reduced asynchronies smaller than 19 ms over the original performance and the recordings with fully increased asynchronies. This finding suggests that listeners might prefer performances containing asynchronies smaller than the perceptual threshold, but with natural timing variabilities given by the ensemble playing, which might make the performance distinguishable from that generated by a computer.

Research has also investigated the roles of music training on perceptual sensitivity to synchrony, providing complex results. As anticipated above, Goebl (2003) found that musically untrained listeners were not able to discriminate the temporal order of pure tones, complex harmonic, or acoustic piano tones; Wing, Endo, Yates, et al. (2014) reported that non-experts were sensitive to the degree of asynchrony when judging the level of togetherness in string quartet performances. Hofmann, Wesolowski, and Goebl (2017) observed no effect of the expertise of the listeners (i.e., musicians, non-musicians, dancers) when rating their preferred jazz-trio performance from a pool of stimuli with manipulated asynchronies. More recently, it has been shown that the ability to detect asynchrony and discriminate temporal order at onsets can increase with multi-hour training in asynchrony and order tasks, and even a single exposure to such perceptual tasks can yield learning (Mossbridge et al. 2006).

In summary, although asynchronies during music performances are inevitable and, to some extent, desirable, they are not always perceived by listeners. A number of factors can contribute to the detection of asynchronies, such as differences in frequency, duration, and sound intensity between two pure tones. The perceptual threshold is likely to increase with the complex onset behaviours of acoustic music events such as acoustic piano tones, and might change with experience of the listeners.

This section has reviewed the literature investigating the perception of synchronization. The following section reviews the development of the methods used for the study of synchronization in ensembles as manifested in the sound produced, highlighting their limitations when applied to the study of ensemble singing. An alternative method, largely applied in the context of solo singing, which has the potential to allow the reliable measurement of synchronization in singing ensembles, is also described.

2.5 Methods to Measure Synchronization

2.5.1 Acoustic Analysis

The study of synchronization is intrinsically a quantitative science. When applied to the analysis of ensemble performances, it often involves the collection, processing and interpretation of large amounts of data. The complexities in data collection and extraction, among other factors, have denied the quantitative analysis of music performances any practical applications until the advent of personal computers, technological advances in signal processing (de Cheveigné and Kawahara 2002; Ifeachor and Jervis 2002) and the development of new methods for signal analysis in the late 1970s. These technological advances boosted a body of work investigating music performances (Devaney et al. 2011), which also included the study of interpersonal synchronization within music ensembles pioneered by Rudolf Rasch (Rasch 1979; Rasch 1988).

Nevertheless, the concept of objective and quantitative analysis of music performance was in place from the late 19th century, when the advent of acoustic analysis radically changed approaches to the understanding of music performance, introducing quantitative methods to measure objectively voice quality, singing and instrumental performances. The scientific approach to music performance was mostly pioneered by the work of Dayton Miller, chair of the Department of Physics at the Case School of Applied Science in Cleveland, and by the body of research conducted by music psychologist Carl Seashore and his colleagues at the University of Iowa. Although Miller's early studies at the beginning of the 1890s involved mainly qualitative analyses of sound, due to the limitations of the equipment available, he was later able to record and analyse quantitatively the fundamental frequency and harmonics from different musical instruments (e.g., flutes, clarinets, violins, saxophones, oboes, and organ pipes). He made use of his newly invented phonodeik (a sound recording apparatus converting sound waves in visual images) in 1908, and the Olaus Henrici's harmonic analyser to perform a frequency-domain analysis of music performances (D. Miller 1916). Specifically, by graphically representing the sound waves, the phonodeik allowed Miller to evaluate the sound patterns through the visualization of the pitch information before the invention of the electronic oscilloscope.

Carl Seashore and his collaborators at the University of Iowa also envisioned the application of scientific tools, methods and procedures to the understanding, evaluation and training of the musical mind. They believed that science could be productively applied to music psychology, pedagogy and performance practices, providing useful information in the musical field. He believed that science:

> *[…] gives us a psychology of music in that it furnishes describable and verifiable facts as a basis for classification […] It furnishes us a technique for the development of musical esthetics […] It forms a basis for the analysis and evaluation of musical talent and will furnish helpful data for vocational and avocational guidance in music […] It lays the foundation for musical criticism, musical biography and autobiography, and musical theory in general […] It furnishes the foundation for the essential facts for the construction of the curriculum, for the selection and motivation of the musically educable, for the evaluation of progress in training, and for countless improvements in the technique and economy of teaching […] It helps to give music its true place* (Seashore 1938, 12)

Upon this belief, Seashore and his colleagues conducted a large number of studies to evaluate objectively music performances, mainly piano, violin and singing performances, through the revolutionary application of acoustic analysis (Seashore 1936; Seashore 1938; Seashore 1947). A number of acoustic parameters were measured, such as intonation of accompanied singing performances of Gounod's *Ave Maria* (Schoen 1922), gliding between notes in recordings of Gounod's *Ave Maria* as well as Handel's *Messiah* (Seashore 1936), and the characteristic attributes of vibrato (Easley 1932; R. Miller 1936; Bartholomew 1934; Metfessel 1932; Seashore 1938), which represents a periodic oscillation of the fundamental frequency (f_o) (Dejonckere, Hirano, and Sundberg 1995).

Seashore and his collaborators took full advantage of the improvements of existing tools, such as microphones, the high-speed oscillograph and the harmonic analyzer for timbral investigation, the stroboscope, a type of oscillograph for the estimation of the f_o (Metfessel 1929), and the voltmeter to measure sound intensity (Tiffin 1932). Their work was also conducted through various new scientific instruments invented by Seashore, such as the audiometer, the tonoscope, the spark chronoscope, the sound-perimeter, the Iowa psychergograph, tone generators and phonophotographic cameras (Miles 1956). Among others, the tonoscope and the phonophotographic apparati were largely used in Seashore's laboratory to evaluate singing. The former allowed the analysis of vibrato from f_o extraction (Seashore 1914; Seashore 1929), while the latter recorded frequency, duration, and intensity on so-called performance scores related to the perceptual measures of pitch, timing, loudness, timbre and emotional expression (Seashore 1938).

Due to the limitations of the technologies available at that time, the data collection and the manual processing work proved to be extremely laborious. This in turn impacted the studies' design and the range of music aspects under investigation. Only short performances of very few musicians were investigated, and the study of ensemble performance was never systematically approached. Finally, the complexities regarding the data collection and analysis affected the interest in empirical performance analysis, which dropped during the 1950s and 1970s. Few objective studies were conducted during those decades, mainly focused on the analysis of different types of voice of an operatic baritone (Fry and Manén 1957), and on the study of amplitude and intensity of the singing voice (Sacerdote 1957).

Despite the extensive work conducted by Seashore and colleagues, their investigations represent early speculative descriptions of some specific aspects of music performance. The generalizable implication of their conclusions is limited due to the small sample sizes, and the lack of statistical analysis, sophisticated digital recording instruments, and music performance theory models available at the time of the research. Nevertheless, the early empirical interdisciplinary explorations conducted in the first half of the 20th century laid the foundations for later studies, by providing the principles and the fundamentals for more robust work conducted in the late 1970s (Devaney et al. 2011). This new corpus of research also included early studies measuring interpersonal synchronization during ensemble performances in three professional trios, a recorder trio, a reed trio, and a string trio, and reporting asynchronies in the order of tens of milliseconds (Rasch 1979; Rasch 1988).

As anticipated above, the resurgence of an active interest in music performance studies in the 1980s coincided with the advent of personal computers, which replaced the slow, expensive and limited machines with smaller, faster and more affordable units that allowed easy recording, collection, and analysis of the properties of the sound. Furthermore, the development of new methods for signal analysis, such as the YIN algorithm for the estimation of f_0 of speech and musical sounds (de Cheveigné and Kawahara 2002; Markel 1972), evolutions in digital signal processing (Ifeachor and Jervis 2002), and the development of computer programs for the analysis of voice (e.g., Praat, SONIC Visualizer, RTSPECT and AMPACT) have made instrumental and singing analysis easily accessible to the general public. The renewed interest in music performance science was also due to a new trend that could be observed in the late 1970s across Europe: an increasing number of professionals had a university degree in the humanities or music and also in physics or engineering, giving them competencies in science and art as two complementary disciplines (Widholm 2012). Various active centres in Music Acoustics flourished, such as KTH in Stockholm led by Johan Sundberg, IRCAM in Paris, the CIRMMT at McGill University in Montreal, CNRS in Marseille, VUZORT in Praha, and the Institute of Music Acoustics *Wiener Klangstil* at the University of Music and Performing Arts in Vienna, which applied quantitative methods to the study of music performance.

The resulting renewed interest in the scientific approach to music performance analysis has greatly increased our understanding of music performance, and has boosted a corpus of research analysing interpersonal synchronization in ensembles. As presented above, some distinctive features of synchronization within and between cultures have been underscored as well as some external factors (e.g., visual contact and leadership) affecting synchronization between musicians in instrumental ensembles, through the successful application of acoustic analysis.

The focus on the analysis of synchronization in instrumental ensembles, with an increasing number of studies using digital pianos to analyse piano duos (Kawase 2014; Bishop and Goebl 2017; Bishop and Goebl 2015; Bishop and Goebl 2018; Ragert, Schroeder, and Keller 2013; Goebl and Palmer 2009), and the lack of investigations in the context of ensemble singing is largely due to the limitations of the current techniques. The successful application of acoustic analysis to the study of ensemble performance is, in fact, largely dependent on the type of music performance examined. To analyse sound signals recorded simultaneously, it is fundamental that each signal is separated. This is possible with digital instruments, such as computer-based pianos, where the signal can be recorded directly at its source. However, this separation has not yet been optimized in the context of ensemble singing, where bleed from the other singers cannot be eliminated completely.

Recently, there have been a number of attempts to analyse synchronization in singing duo performances (Palmer et al. 2013) and also in string quartets (Timmers, Endo, and Wing 2013; Timmers et al. 2014), in which the sound is not digital at its source, using close proximity microphones and algorithms designed *ad-hoc* for the studies. There have also been attempts to investigate singing ensemble performances, with tuning as the major topic of interest, through a machine-learning approach applied to audio recordings (Devaney, Mandel, and Fujinaga 2012). Although close proximity microphones capture the data of the individual singers, they do not eliminate completely the bleed from other performers (Daffern and Brereton 2013). The difficulties in isolating individual notes in ensemble singing remain, and acoustic analysis does not, at present, represent a fully reliable method for the analysis of ensemble singing.

It is important to acknowledge that the application of acoustic analysis in the science of singing has improved our understanding of a number of aspects of solo singing, such as tuning (Devaney et al. 2011; Sundberg, Prame, and Iwarsson 1995), vibrato (Prame 1994; de Almeida Bezerra et al. 2009), vocal registers (Hirano, Hibi, and Sanada 1989; Morris et al. 2012), formants (Sundberg 1987; Sundberg 2001), singer *versus* non-singer voices (Brown et al. 1993; Åkerlund, Gramming, and Sundberg 1992), singing styles (Thalén and Sundberg 2001; Manfredi et al. 2015), voice development from childhood to adulthood (Barlow, Lovetri, and Howard 2007; Welch and White 1993), perceptual variables (Kenny and Mitchell 2006; Kenny and Mitchell 2007), and real-time visual feedback (Welch, Howard, and Rush 1989; Howard et al. 2007). However, in spite of these important contributions to the study of solo singing, acoustic analysis does not constitute a fully reliable technique for the study of synchronization between singers in singing ensembles.

2.5.2 Electrolaryngography – Electroglottography[1]

The complexities of polyphonic analysis associated with audio recordings were recently avoided by David Howard, analysing tuning in two different soprano-alto-tenor-bass (SATB) ensembles through the application of acoustic analysis in conjunction with electrolaryngography (Lx) to extract the f_0 estimates from vocal fold contact information, (Howard 2007b; Howard 2007a; Howard 2007c). Electrolaryngography and electroglottography (EGG), two non-invasive techniques that assess vocal fold vibration in vivo through electrodes placed externally on either side of the neck at the level of the larynx, allow measurement of performance data in solo and ensemble singing performances. Based on the principle that human tissue is a good electrical conductor while air is not, the Lx/EGG electrodes monitor the closing and opening of the vocal folds (see **Figure 2.1**), by measuring variations in the electrical impedance of the larynx. A constant, high frequency current in the 0.3-3 MHz range is sent through the neck of the participant using the electrodes: the impedance/admittance variation of the current, caused by the contacting and de-contacting of the vocal folds, is measured (Fourcin and Abberton 1971), which reflects the amount of contact between the vocal folds (Scherer, Druker, and Titze 1988) and is graphically represented in the waveform produced.

---

[1] Most part of this section has been published in D'Amario and Daffern (2017). Copyright 2017 by American Psychology Association. Adapted with permission.

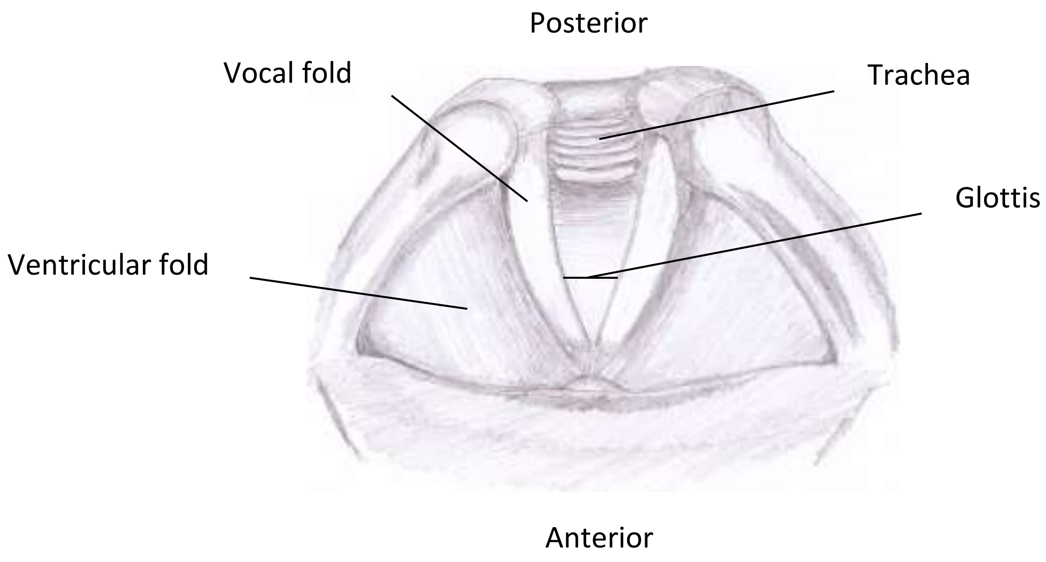


Figure 2.1 *Schematic representation of the vocal folds*[2] *in open position, which are located within the larynx at the top of the trachea. The gap between them is called the glottis. Figure from D'Amario and Daffern (2017). Copyright 2017 by American Psychology Association. Adapted with permission.*

The Lx/EGG waveform allows the objective measurement of a number of aspects of vocal fold vibration occurring during phonation[3]. The most common quantitative parameters include:

- Amplitude (Amp): peak-to-peak amplitude of each cycle.
- Fundamental period (T_0): duration of one vibratory cycle.
- Fundamental frequency (f_o): calculated as $f_o = \frac{1}{T_0}$.
- Vocal fold closed phase (CP): the time in each cycle for which the vocal folds remain in contact.
- Vocal fold open phase (OP): the time in each cycle for which the vocal folds remain apart.
- Vocal fold contact quotient (CQ): an estimation of the time for which the vocal folds are in contact. This parameter was also referred to as "closed quotient" (Fourcin and Abberton 1971; Howard 1995) or as "quasi closed-quotient" (Hacki 1996);

---

[2] Vocal folds are constituted by muscles shaped as folds and are approximately 9 to 13 mm and 15 to 20 mm long in adult females and males, respectively (Sundberg, 1987, 6).

[3] '*Phonation* means sound generation by means of vocal fold vibration' (Sundberg 1987, 9)

nevertheless, the term "contact quotient", introduced by (Davies et al. 1986) in relation to EGG, recognises that it does not reflect a full glottal closure (Herbst et al. 2017).

- Vocal fold open quotient (OQ): an estimation of the time for which the vocal folds remain apart.

Lx/EGG is a commonly used tool for the analysis of the laryngeal vibratory system that governs the singing voice, improving our understanding of a number topics, comprising different singing styles, registers, voice classifications, tuning, resonance strategies, vibrato, vocal development, practice effects, voice quality, analysis techniques, and its use in clinical research and vocal pedagogy (for a recent systematic literature review of evidence-based studies using Lx/EGG to assess the above aspects see D'Amario and Daffern 2017). Lx/EGG combined with other methods, such as acoustic analysis, electrophotoglottography, electromyography, inverse filtering, measures of sub- and supra-glottal pressure and airflow, have advanced knowledge of the physiological mechanisms underpinning the singing voice, revealing the contribution of the voice source. These research techniques and knowledge can usefully inform traditional psychomusicological approaches to singing performance, to better grasp the intentions, priorities, and perceptions of the performers as well as the audience. The wealth of literature emerging from Lx/EGG research is highly valuable and leads toward a fuller understanding of the singing voice.

Despite these advances, the application of Lx/EGG has mainly focused on the analysis of solo singing, with only very few studies analysing singing ensembles, and always with tuning as the main factor under investigation (Howard 2007b; Howard 2007a; Howard 2007c). Caution is recommended for the characterisation of Lx/EGG waveshape in ensemble singing, since a recent study conducted by Ternström, D'Amario, and Selamtzis (2018) has shown among other aspects that EGG waveshape is highly variable across professional singers and across the voice range profile. On the other hand, the successful application of Lx/EGG in combination with acoustic analysis in the context of polyphonic recordings suggests that the combined application might be a useful tool to investigate other aspects of music ensemble performances that require tracking of the individual voices, such as synchronization in singing ensembles. The possibility to track the f_0 estimates from Lx/EGG recordings of a single singer in a polyphonic context makes these techniques potentially highly beneficial for the analysis of interpersonal synchronization in ensemble singing. Further investigations testing the use of Lx/EGG for the study of synchronization in ensemble singing are needed to evaluate its effectiveness in this field.

In addition to the collection of performance data and the identification of individual contributions during ensemble performances discussed above, the study of interpersonal synchronization relies on the detection of note onsets and offsets, which then enables the objective measurement of temporal coordination. The most common methods of onset/offset detection are reviewed in the next section.

2.5.3 Onset/Offset Detection

One common method for onset/offset detection consists of the manual annotation of note onsets/offsets from the audio waveform, using specific software for the analysis of audio recordings, such as Sonic Visualiser (Cannam, Landone, and Sandler 2010) and Praat (Boersma 2001; Boersma and Weenink 2013). For example, note onsets of some violin performances have been manually identified by six musically-trained judges using Sonic Visualizer to investigate interpersonal synchronization in piano-violin duo performances (Bishop and Goebl 2015). However, the successful application of this method is associated with the type of music recording. This manual annotation is suitable only in the case of percussive and pitched instruments, where the note onsets-offsets are easy to detect in the audio waveform. In addition, this approach is highly time-consuming and dependent on the subjective evaluation of the experts manually analysing the recordings (Goebl, Dixon, and Schubert 2014).

To address much of this subjectivity and allow investigation of larger-scale studies, several algorithms have been suggested that automatically detect note-onsets from audio recordings (for a review, see Bello et al. 2005). A few studies have focused on spectral features of the signals (Masri 1996), combined phase and energy information (Bello et al. 2004), analysed phase deviations across the frequency domain (Bello and Sandler 2003), considered change of energy in frequency sub-bands (Klapuri 1999), or are based on probabilistic methods such as hidden Markov models (Abdallah and Plumbley 2003). Other approaches are based on the fundamental frequency contour and sound level envelope (Friberg, Schoonderwaldt, and Juslin 2007), or on time and frequency domain features (Dixon 2001). The selection and reliability of the algorithms mentioned above are strictly correlated to the type and quality of the audio signal; for example, time domain methods perform relatively well if the signal is very percussive as in piano or drum recordings.

It is noteworthy that existing algorithms perform less well in singing compared with other classes, such as solo brass, wind instruments, and polyphonic pitched instruments. Whilst onsets and offsets are often clearly distinguishable for percussive sounds, in singing these are very clear only if a noticeable pause between notes is provided. In singing, within a legato phrase, onsets and offsets vary according to vibrato, vocal fluctuations, timbral characteristics, and onset envelopes, especially where consonants are absent.

In Music Information Retrieval Evaluation eXchange (MIREX 2016), the best-performing algorithm for the onset detection of solo singing voice achieved an F-measure (a metric of the overall performance) of 61.7 %; whereas, the best-performing algorithms for drums, plucked strings, brass and wind instruments achieved F-measures of 93 %, 92 %, 91 %, and 78 %, respectively. Toh, Zhang, and Wang (2008) implemented a system for the analysis of the solo singing voice that accurately identified 85 % of onsets within 50 ms of the ground truth, i.e. the manually annotated values of the same recordings. However, when considering interpersonal synchronization, this is not precise enough for the analysis of the highly accurate coordination that is found in professional music ensembles, known to be in the order of tens of milliseconds (Rasch 1979; Rasch 1988; Keller 2014). An algorithm detecting onsets/offsets in singing performances with a tight temporal resolution would be beneficial to improve our understanding of interpersonal synchronization in ensemble singing. To complicate this matter further, synchronization in the context of ensemble singing might be affected by the lyrics, as the duration of consonants might change based on the phonetic context (Sundberg and Bauer-Huppmann 2007). Research has shown that in several Lieder selected from Robert Schuman’s *Frauenliebe und leben*, op. 42 and *Dichterliebe*, op. 48, the tone onset of piano accompaniment is often synchronized with the singers’ vowel onsets, although researchers found exceptions to this principle due to expressive reasons (Sundberg and Bauer-Huppmann 2007).

Once onsets have been detected, the next and final step required for the study of synchronization involves the interpretation of the data extracted. Different techniques have been implemented, based on the type of recordings collected, that allow the evaluation of interpersonal synchronization in music ensembles. These techniques are presented in the following section.

2.5.4 Statistical Techniques

The most common method used for the evaluation of interpersonal synchronization comprises the analysis of the absolute asynchrony, which represents an absolute matrix of the temporal lag between two sound events. This is calculated subtracting the time of one event from that of another; for example, subtracting the onset of a soprano from that of a mezzo-soprano. Linear statistical techniques such as signed and absolute mean, median, standard deviation (SD), and coefficient of variation (CV) have then been frequently applied to raw data to investigate the precision and consistency of the synchronization during ensemble performances. Analysis of variance (ANOVA) has then been used to investigate the effects of particular variables on asynchrony values (Keller and Appel 2010; Goebl and Palmer 2009; Loehr and Palmer 2009). More recently, multilevel linear modelling (Gelman and Hill 2007), which allows the evaluation of fixed and random effects, have also been applied to evaluate the effects of multiple, nested variables on interpersonal synchronization (Bishop and Goebl 2014; Bishop and Goebl 2015).

Circular statistics might also be applied, computing the relative phase of each point in the time series relative to another, so the temporal difference is expressed in respect to the time cycle or period of one variable (Mardia and Jupp 2000). Circular statistics measures commonly used in this field of research are the mean vector, quantifying the mean phase angle [μ], and the degree of entrainment or synchronization (r, on a scale from 0 to 1). Such measures are very appropriate when the relative phase of two time series is significantly far from 0 degrees, as in the case of most ethnomusicological studies analysing Indian, Afro-Brazilian Congado, and Transylvanian music traditions (Clayton 2015; Lucas, Clayton, and Leante 2011; Clayton 2007b; Bonini-Baraldi, Bigand, and Pozzo 2015). It has been shown, in fact, that in such cases the linear statistics do not provide information about relevant synchronization patterns (Clayton, Sager, and Will 2005).

On the other hand, if the phase difference/mean asynchrony is not far from 0 and the two rhythms are strongly entrained/synchronized, as in the case of most Western Classical ensemble traditions, the two approaches offer different perspectives but yield similar patterns of findings (Hove, Keller, and Krumhansl 2007). Still, the use of asynchrony calculations rather than relative phases has the advantage of avoiding arbitrary decisions on what the period is, which might represent an issue in some highly metrical pieces. The identification of the period, in fact, might not be straightforward in case of a high hierarchical metrical structure of the piece; the selection of one period or the other would, ultimately, provide give different phase calculations. For this reason, asynchrony computation represents the most appropriate approach to the analysis of interpersonal synchronization in the Western Classical repertoire.

2.5.5 Mixed-Methods Approaches

In addition to the methods described above and focused on the analysis of synchronization as manifested in the sound produced, research has also made use of mixed techniques when investigating simultaneously the synchronization of onsets/offsets and body movements. For instance, in parallel with acoustic analysis, research has used systems tracking movements in vivo, such as Motion Capture (MoCap) and Microsoft Kinect, and computer vision techniques (for a review, see Jakubowski et al. 2017), such as dense optical flow, estimating body movements from video recordings to investigate the body movements between performers (Badino et al. 2014; Glowinski et al. 2013; Glowinski et al. 2012; Eerola et al. 2018; Jakubowski et al. 2017). More recently, it has been shown that motion capture technologies can be particularly useful for the analysis of synchronization in repertoires of oral tradition such as the Transylvanian village music, in which written instructions are not available (Bonini-Baraldi, Bigand, and Pozzo 2015). A simultaneous analysis of aspects of onsets/offsets synchronization and body coordination may be useful to contextualise the findings related to sound synchronization revealing the aspects of body movements that underpin such synchronization. However, the analysis of body movements is beyond the scope of this thesis, and thus these technologies are not reviewed in detail here.

A limited number of studies have also applied other techniques to investigate the synchronization of musicians' physiological data during ensemble playing. For example, Babiloni et al. (2011) used electroencephalographic (EEG), electrooculographic (EOG), electromyographic (EMG) techniques and acoustic analysis to investigate relationships between physiological data in four saxophonists playing in an ensemble. Müller and Lindenberger (2011) analysed the synchronization of cardiac and breathing patterns between singers during choir singing. Synchronization of physiological data between musicians is also beyond the scope of this project, in which the focus is on the methods associated with the analysis of the music recordings and the singing voice, i.e. acoustic analysis and electrolarygography/electroglottography.

Another valuable technique might be the study of verbal interactions during rehearsals, in addition to the objective measurements of synchronization. The analysis of verbal interactions between musicians is commonly used to investigate the rehearsal strategies commonly used during rehearsal to excel ensemble performance (Ginsborg, Chaffin, and Nicholson 2006; Ginsborg and King 2007; Ginsborg and King 2012). An approach comprising both the analysis of verbal discussions and interpersonal synchronization might provide an important contribution to the field of temporal coordination, revealing the deliberate rehearsal strategies that musicians apply to enhance synchronization and excel music performance.

2.5.6 Summary

In summary, this section described the current techniques for the analysis of music recordings and the study of interpersonal synchronization, highlighting the complexities of timing detection in non-percussive and polyphonic recordings, such as singing ensemble performances. A robust method for onset detection in non-percussive recordings still needs to be designed and evaluated.

2.6 Conclusions

Interpersonal synchronization in musical ensembles is a fundamental performance parameter that contributes to the expressiveness of ensemble performances. It is maintained through iterative temporal adaptations with co-performers, and supported by complex and highly connected, cognitive, motor, and (neuro)physiological mechanisms, including brain activity, heart and respiration synchronization, and coordination of musicians' body movements. A range of factors might influence interpersonal synchronization, such as the visual contact between singers, leader-follower relationships, and rehearsal. However, the nature of these effects is not currently fully understood. Most research has focused on the study of instrumental ensembles, whilst the analysis of interpersonal synchronization in ensemble singing currently lacks thorough investigation. Synchronization between singers during singing performances is unique, as singers only rely on their body to sing, i.e. a singer's body is their instrument. The addition of text during singing performances also makes ensemble singing an interesting context for the study of synchronization, as singers do synchronize vowels, consonants and syllables, with the co-performer(s) during ensemble performances. Although asynchronies are very common even in professional ensembles, they are not always perceivable, due to a number of factors, such as the complexities of onset characteristics of real music events and the experience of the listeners. Acoustic analysis represents a common and successful method for the evaluation of synchronization in instrumental ensembles. The analysis of synchronization in the context of ensemble singing is particularly difficult, due to issues regarding the identification of each singer's individual contribution in polyphonic recordings, and the problems arising from the detection of timing information in non-percussive instruments, in which the identification of the "true" onset/offset is highly debatable. A new method for onset/offset detection in singing and for the analysis of interpersonal synchronization in ensemble singing still needs to be identified and tested. Such methods might provide a novel and profound contribution to the study of interpersonal synchronization in ensembles.

3 A New Algorithm for Timing Detection in Ensemble Singing

3.1 Introduction

As discussed in the previous chapter, the analysis of interpersonal synchronization in ensemble singing represents a major challenge due to difficulties of extracting individual voices from ensemble recordings, and identifying true onsets and offsets that can vary in relation to vibrato, timbre, and onset envelopes, especially in the context of a legato phrase. Currently, there is no a robust algorithm able to extract timing information related to onsets/offsets in the context of ensemble singing nor individual voices. The automatic detection of timing information would be highly beneficial for the investigation of interpersonal synchronization in singing ensembles as well as the study of other aspects of sound analysis, including music information retrieval and music transcription.

This chapter[4] addresses the complexities of analysing onset and offset timings in polyphonic singing recordings through a novel algorithm, termed TIMEX, that automatically extracts timing information based on the combined application of electrolaryngography (Lx) and acoustic analysis. The combination of these two tools proved to be a successful method to identify the individual contribution of each singer during ensemble singing as demonstrated in previous studies analysing intonation in soprano-alto-tenor-bass (SATB) quartets from fundamental frequency (f_o) estimates (Howard 2007a; Howard 2007b; Howard 2007c). Based on such evidence, it was hypothesized that this combined application is a valuable tool for the onset/offset detection in polyphonic singing recordings, fundamental to the analysis of synchronization in ensemble singing by tracking the f_o profile of the individual voices.

The remainder of this chapter is organized in four sections. Firstly, an overview of the novel algorithm for timing extraction is presented, then the algorithm is tested on a set of singing duo recordings. Finally, results are discussed, and conclusions presented.

---

[4] Most of this chapter, including figures and tables, is reported in D'Amario, Daffern, and Bailes (2018b), and is © the authors, licensed CC-BY.

3.2 Algorithm Description

In singing, as in instrumental performance, the horizontal degree of note overlapping and separation of successive notes (i.e., articulation) varies for expressive reasons, providing the staccato or legato characteristics of the phrase/piece. When performing staccato, the successive sounds are very well separated. When singers perform legato and to a vowel or in cases when no unvoiced consonants are present, there are no silences between the notes within a phrase: phonation continues until the next rest/breath, effectively creating a *portamento* (i.e., gliding) between notes. In light of the different types of articulation of singing performances, it was therefore necessary, in the development of the algorithm, to set criteria with which to analyse the beginning and ending of each note within the piece. This has resulted in four categories being defined to denote the true beginning and ending of the scored notes. These are shown in **Figure 3.1** and defined as:

- Onset (ON): beginning of phonation after a silence.
- Note ending (NE): peak/trough in f_0 during phonation within a legato phrase, that is atypical of a vibrato cycle's characteristics for extent and frequency, calculated between 80 and 120 cents and between 2 and 9 Hz, respectively, and refined for each singer.
- Note beginning (NB): peak/trough in f_0 during phonation that exceeds the maximum vibrato extent and is less than the vibrato frequency following a note ending.
- Offset (OF): ending of phonation followed by a silence.

In order to automate the extraction of the above categories, the following definitions have been formulated and parameter values inputted. The values were manually determined by testing with several recordings and can be modified by users.

- *Break*: a sequence of one or more points where the Lx signal is null.
- *Rest*: a sequence of a minimum number of consecutive points where the Lx signal is null. The number of minimum points required to classify a break as a rest is arbitrarily defined; for this specific set of recordings, it has been set to a corresponding time window of 300 ms representing a quaver rest at 100 beats per minute (BPM). This parameter has been specifically formulated to allow the identification of ONs and OFs after and before a rest, respectively.

- *Phrase*: a section of the Lx recording comprised between an ON and the following OF.
- *Fluctuation:* the difference in frequency between two Lx or audio points; the fluctuation can be linear or logarithmic, depending how it is measured. For these recordings, it was set to 80 cents, representing a fluctuation smaller than a semitone in the equal temperament system, which is equal to 100 cents.
- *Local max*: a point where the Lx/audio value is higher than the Lx/audio values at the previous and at the following point.
- *Local min*: a point where the Lx/audio value is lower than the Lx/audio values at the previous and at the following point.
- *ON/OF fluctuation range*: the range of points after an ON or before an OF, where the singer’s voice typically oscillates; local max/min points are ignored within this range, because they are not aligned with note changes, but are the result of the vibrato. Its duration is arbitrarily defined. For this specific set of recordings, a value of 300 ms has been set corresponding to a quaver note at 100 BPM, representing half the value of the shortest note onset (note 19) and offset (note 7) of this piece.
- *Vibrato frequency threshold*: the minimum frequency of oscillation of the Lx or audio signal that classifies the segment as vibrato, and therefore is not associated with a true note change from the score. For these recordings, it was set to 5 Hz.
- *Local peak*: a point with a positive Lx value that falls in the middle of a range of a prescribed temporal window, where at least one point with null Lx frequency exists before and after such a point. The temporal window to conduct the check is arbitrarily defined; a time span of 500 ms (equal to about a dotted quaver rest at 100 BPM) placed around the point in question has been used with satisfactory results in this project.
- *Spiking range:* a range of points immediately before an ON or after an OF, where the Lx signal artificially spikes relative to the corresponding audio signal. The width of such a range is arbitrarily defined; given the steepness of the spikes, a value of just 10 ms has proven sufficient to isolate the spikes.

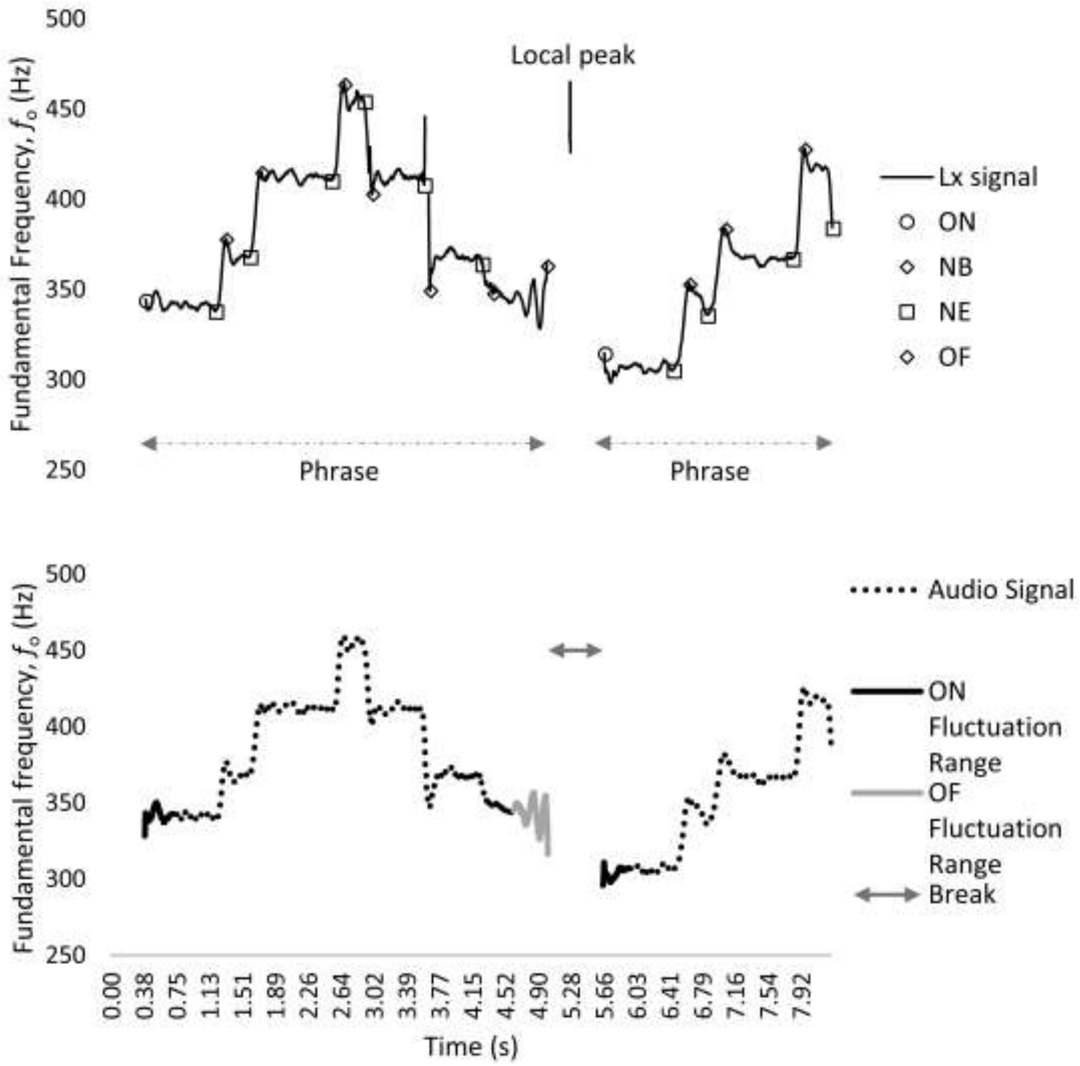


Figure 3.1 *The f_o profile of measures 1-3 of the raw Lx and audio signal from an upper voice performance of the two-part piece composed for this study, showing: i) on the top panel, the Lx recording with the four sets of categories identified for each note within a legato phrase [i.e., onset (ON), note beginning (NB), note ending (NE), and offset (OF)], a local peak and two phrases; ii) on the bottom, the audio recording, with the ON and OF fluctuation ranges and the break range.*

TIMEX detects and extracts ON, NB, NE, and OF ensuring consistency of the analysis based on the following steps, as shown in **Figure 3.2**.

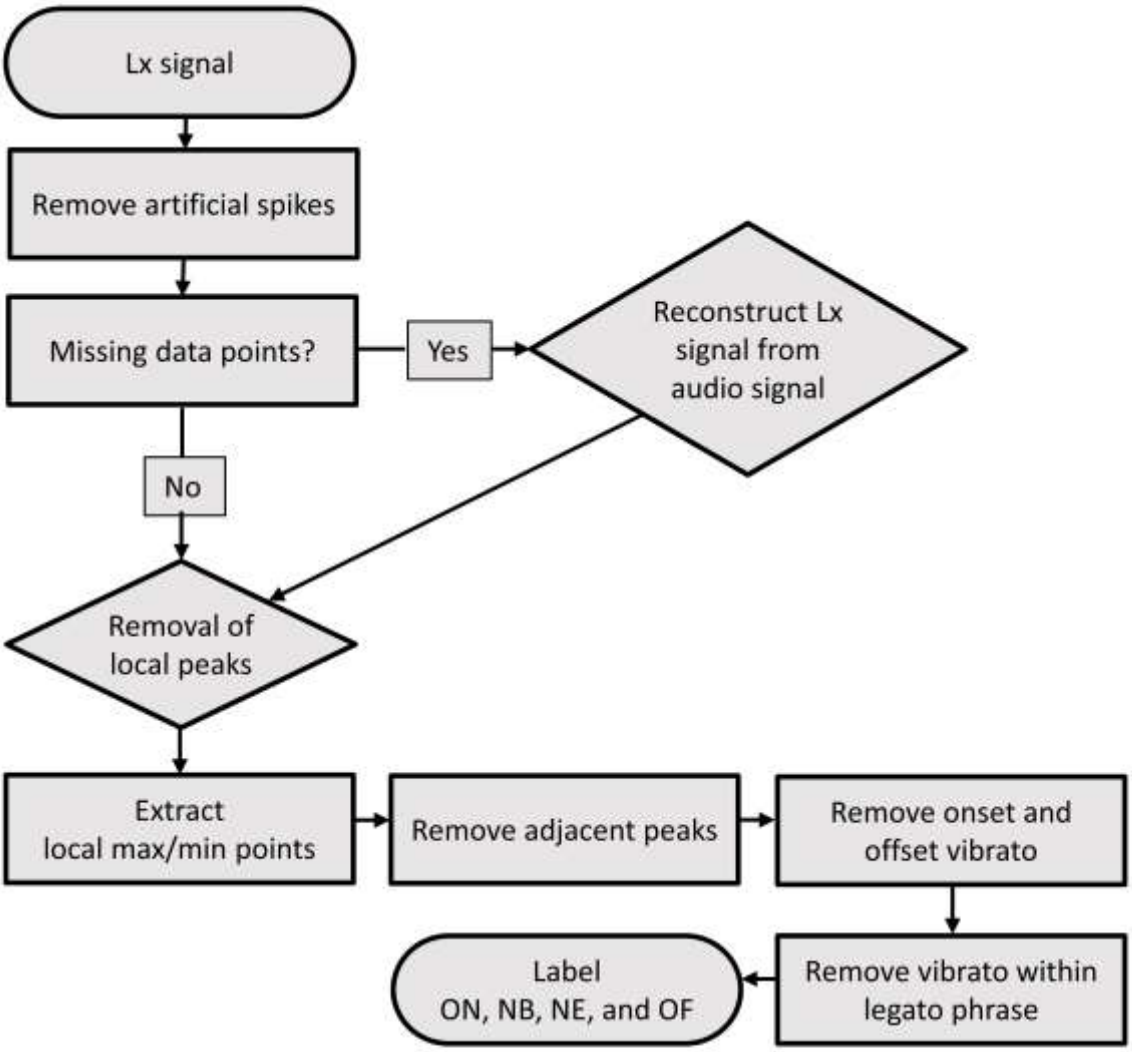


Figure 3.2 *Algorithm flowchart.*

Step 1: removal of Lx readings in the spiking range. The first operation performed on the raw Lx data is to remove all the positive Lx readings within the spiking range (adjacent to the breaks), replacing them with null values. This step is executed to prevent the artificial spikes from leading to a skewed and distorted reconstruction of the Lx signal from the audio signal (the reconstruction procedure is explained in Step 2).

Step 2: reconstruction of the missing Lx signal from the audio signal. If the Lx signal is weak, the algorithm reconstructs the signal from the audio recording. This is achieved through a normalization procedure designed to reconstruct the Lx signal to follow the same shape as the audio signal. The audio signal is scaled to match the original Lx values at the edges of the interval where the Lx signal is missing, therefore avoiding artificial max/min points being generated at the edges; from here on, the “original” Lx signal refers to the signal after the Lx readings in the spiking range have been removed, as per step 1.

Using the following nomenclature:

- t_0, t_1: time intervals at the boundaries of the range where the original Lx signal is missing or weak, and the audio signal is at least partially available.
- f_o_Lx$_0$, f_o_Lx$_1$: the values of the original Lx signal at t_0 and t_1; they are both positive by definition of how t_0 and t_1 are selected.
- f_o_AUDIO$_0$, f_o_AUDIO$_1$: the values of the audio signal at t_0 and t_1; if one of them is zero, it is calculated as the other one multiplied by the ratio between f_o_Lx at that point and f_o_Lx at the other end, while if both are zero reconstruction is not attempted for this interval.
- f_o_Lx$_L$(t), f_o_AUDIO$_L$(t): the values of the linearized Lx signal and the audio signal respectively at time t, with t falling between t_0 and t_1; these are linearized as falling on a straight line connecting f_o_Lx$_0$ and f_o_Lx$_1$, and f_o_AUDIO$_0$ and f_o_AUDIO$_1$ respectively.
- f_o_Lx(t), f_o_AUDIO(t): the values of the original Lx signal and the audio signal respectively at time t, with t falling between t_0 and t_1.

The linearized Lx and audio values are first computed as follows:

$$f_o\mathrm{Lx_L(t)} = f_o\mathrm{Lx_0} + (f_o\_\mathrm{Lx_1} - f_o\_\mathrm{Lx_0}) \cdot \frac{\mathrm{t}-\mathrm{t_0}}{\mathrm{t_1}-\mathrm{t_0}} \tag{1}$$

$$f_o\_\mathrm{AUDIO_L(t)} = f_o\_\mathrm{AUDIO_0} + (f_o\_\mathrm{AUDIO_1} - f_o\_\mathrm{AUDIO_0}) \cdot \frac{\mathrm{t}-\mathrm{t_0}}{\mathrm{t_1}-\mathrm{t_0}} \tag{2}$$

Then, if f_o_AUDIO(t) = 0, f_o_Lx(t) = 0 (reconstruction not possible at a point where even the microphone reading is not available), otherwise f_o_Lx(t) is reconstructed as

$$f_o\_\mathrm{Lx(t)} = f_o\_\mathrm{AUDIO(t)} \cdot \frac{f_o\_\mathrm{Lx_L(t)}}{f_o\_\mathrm{AUDIO_L(t)}} \tag{3}$$

The result of this reconstruction is that the Lx signal follows the shape of the audio signal in the areas where the raw signal is not available, remaining continuous with the original values where present, as shown in the example of **Figure 3.3**.

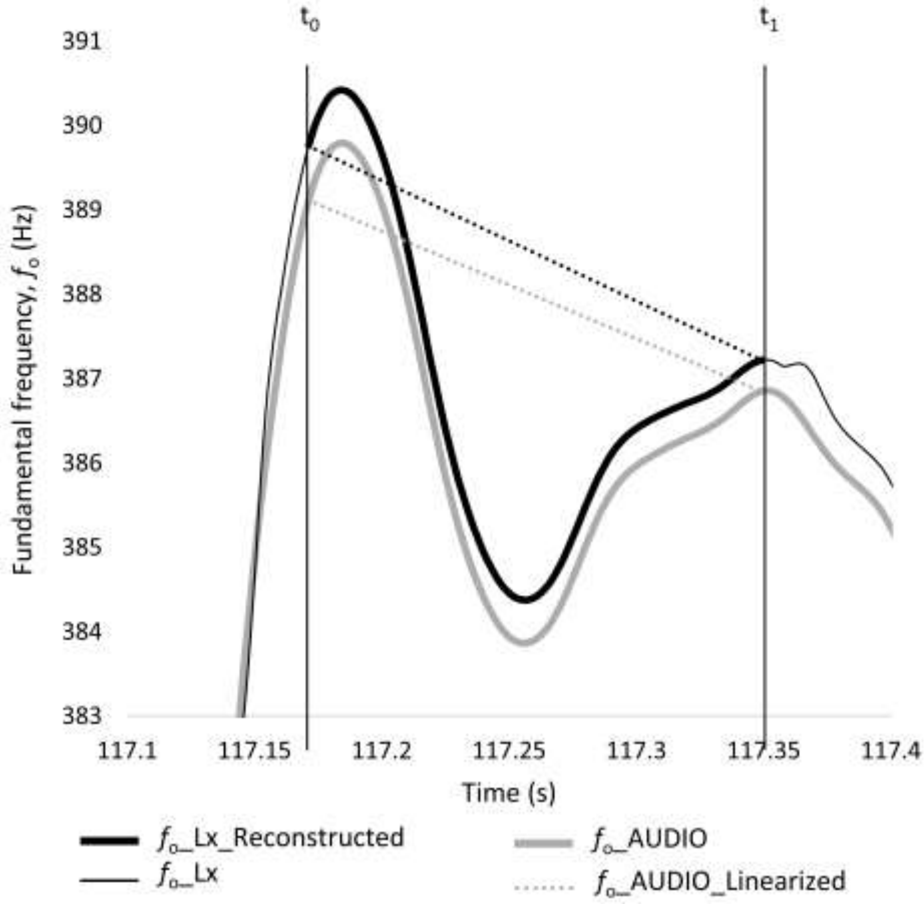


Figure 3.3 *Excerpt of the Lx and audio signals from a recording of the upper voice performance, showing the reconstruction of the f_o_Lx signal from f_o_AUDIO signal in the temporal interval t_0-t_1, in which the Lx signal was missing. The Lx signal was reconstructed (see f_o_Lx_Reconstructed) based on the linearized Lx and audio signal (see f_o_Lx_Linearized and f_o_AUDIO_Linearized, respectively). The small discrepancies in the f_o values between the Lx and audio recordings shown in the figure might be due to a number of factors, such as the performance of the Praat algorithm used for the f_o extraction, and the fact that the sound was measured at two different points: on the neck using Lx electrodes, and on the cheek at approximately 2.5 cm from the lips through head-mounted microphones, after the sound travelled through the vocal tract. Nevertheless, such discrepancies are not relevant to the timing detection and the synchronization analysis, as these are not affected by the f_o themselves.*

Step 3: removal of Lx local peaks. After the Lx signal has been reconstructed, any remaining local peaks are identified, based on the selected range (see definition above) and removed. The purpose is to eliminate spurious readings that are sometimes produced by the Lx sensor, which typically occur in a narrow time range, and can be identified via a proper selection of the local peak range. Removing the peaks after the signal has been reconstructed, from the audio data where possible, allows the maximum amount of Lx data to be retained. The resulting Lx signal after the removal of the local spikes is defined as the “reconstructed” Lx signal.

Step 4: identification of ONs, OFs, NBs and NEs. Once the Lx signal has been reconstructed, it is processed to extract ONs and OFs of phonation and local max/min points during phonation. Then, local max/min points are retained if all the following conditions are satisfied:

1. The point is not too close to the adjacent local max/mins. Points that are too close to each other are removed, to avoid retaining small steps within a tone ascending or descending section as NBs or NEs, when they are just fluctuations of the singer’s voice that sometimes occur within a note change. A value of just 10 ms is sufficient to discriminate those points from the max/mins to be retained.
2. The point does not fall within the ON or the OF fluctuation range.
3. Any of the following two conditions are satisfied:

   3.1. The logarithmic fluctuation, measured in cents, of the current point from the previous onset or max/min, or to the next max/min, is greater than a prescribed threshold. The distance in cents between two points at frequencies f_1 and f_2 is defined as in Howard (2007a)

$$c(f_1, f_2) = 3986.3137 \cdot \log_{10}\left(\frac{\max(f_1,f_2)}{\min(f_1,f_2)}\right) \quad (4)$$

   3.2. The frequency of oscillation of the point, relative to the previous and the next point, is lower than the vibrato frequency threshold; this condition is applied to disregard any max/mins that are the result of a vibrato of the singer’s voice, without having to set a threshold that is too high for the logarithmic fluctuation, which would lead to discarding valid NBs or NEs for semitones. The vibration frequency (vf_o) of the point is defined as the lowest of the oscillation frequencies relative to the previous and the next max/min, as shown in **Figure 3.4**:

$$vf_n = \frac{1}{\max(t_n - t_{n-1}, t_{n+1} - t_n)} \quad (5)$$

The ability to manually tweak the results after visual validation is set to ensure that all and only the relevant max/min points are retained as NBs and NEs.

The algorithm has been designed using the Visual Basic program and run in Microsoft Excel. It reads two sets of data, f_0 estimates in Hertz and corresponding time stamps of the Lx and audio signals in milliseconds with a time step of 1 ms, that need to be first extracted from the recordings and imported in Excel as a tabular list of data, in order for the algorithm to run.

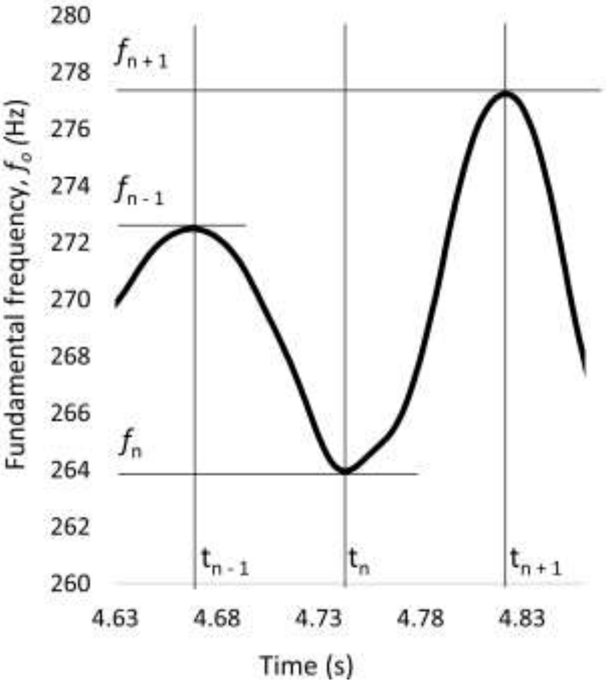


Figure 3.4 *Example of the vibration frequency computed across a full cycle, extracted from an audio clip of the upper voice used for the study.*

3.3 Algorithm Evaluation

In the evaluation of the algorithm, a set of singing duo recordings were used, as described in the following section, and the following three aspects were tested: i) reliability of the Lx signal, ii) effectiveness of the reconstruction process, and iii) performance in the onset/offset detection. These aspects were evaluated through the related tests described in section *3.3.2*, section *3.3.3*, and section *3.3.4* (for an overview of the tests conducted, see **Table 3.1**).

3.3.1 Stimulus

For the evaluation of the algorithm, recordings of a two-part piece (see **Figure 3.5**), composed for the pilot study reported in the following chapter were used. The upper voice has a range of a 7th, whilst the lower voice a range of a 5th; the upper voice features a higher tessitura than the lower voice. The piece comprises eight bars, grouped in four legato phrases, sung legato to the vowel /i/ and separated by rests, thus representing a short, stand-alone piece. The legato was specified to control the articulation, which would likely affect the temporal coordination between singers. Repeated notes were avoided within the legato phrases, because true NBs detection would be very difficult if a noticeable pause in phonation between notes is not produced, as noted in the previous chapter (see section *2.5.3*). Text, consonants and syllables were not included and a vowel was preferred, to control for the potential effect of the lyrics, and syllables on synchronization and simplify the detection of note timing (Sundberg and Bauer-Huppmann 2007), as noted in section *2.5.3*. The choice of the vowel was left to the first duo taking part to the study, who chose the vowel /i/. For consistency, the same vowel was used with the second duo taking part in the pilot, and also with the remaining singing ensembles involved in this thesis (see the singing quintet in chapter *6*).

The number of recordings selected from the entire pool and used for the evaluation changed according to the purpose of the tests. For more information about the characteristics of the piece and the choices upon which the composition of the piece were based, see section *4.2.2*. For more information about the two singing duos performing the piece, see *4.2.1*.

Table 3.1 *Overview of the tests implemented for the evaluation of the new algorithm, including the data sets used for each test.*

Test n	**Test aim**	**Dataset**	**Number and types of recordings**	**Recordings total length**
1	Lx reliability	Entire original	- 96 Lx recordings of the upper voice - 96 Lx recordings of the lower voice - 96 audio recordings of the upper voice - 96 audio recordings of the lower voice	Lx/audio: 83 minutes
2	Reconstruction effectiveness	Subset A	- 20 Lx recordings of the upper voice - 20 Lx recordings of the lower voice - 20 audio recordings of the upper voice - 20 audio recordings of the lower voice	Lx/audio: 20 minutes

3	TIMEX performance	Subset B	- 14 Lx recordings of the upper voice - 14 Lx recordings of the lower voice	12 minutes

3.3.2 Test 1: Lx Signal Reliability

Onset/offset detection based on audio recordings is not fully reliable in the case of singing ensemble recordings as bleed from co-performer(s) cannot be fully eliminated, as explained in the previous chapter. Therefore, the quantification of the percentage of times that the timing detection was not based Lx signals but on the audio recordings is important to evaluate the protocol. The reliability of the Lx signal was quantified through the measurement of the continuous/discontinuous parts of the Lx signal. The analysis of the quality of the Lx signal analysis was conducted on the full set of recordings collected for the pilot study reported in chapter *4*. This comprises 96 Lx recordings of the upper voice and 96 Lx recordings of the lower voice part of the two-part piece, and sung by two duos. Sections of the Lx signal associated with rests in the musical score were not scrutinized, as the Lx was supposed to be null in the absence of phonation. Results show that the Lx signal was unusable for 0.7 % of the recordings, and the algorithm's application to the audio signal was limited to 0.7 % of the full set of audio recordings. Analysis also shows that the discontinuous Lx segments were on average 31 ms long $(SD = 18\, ms)$.

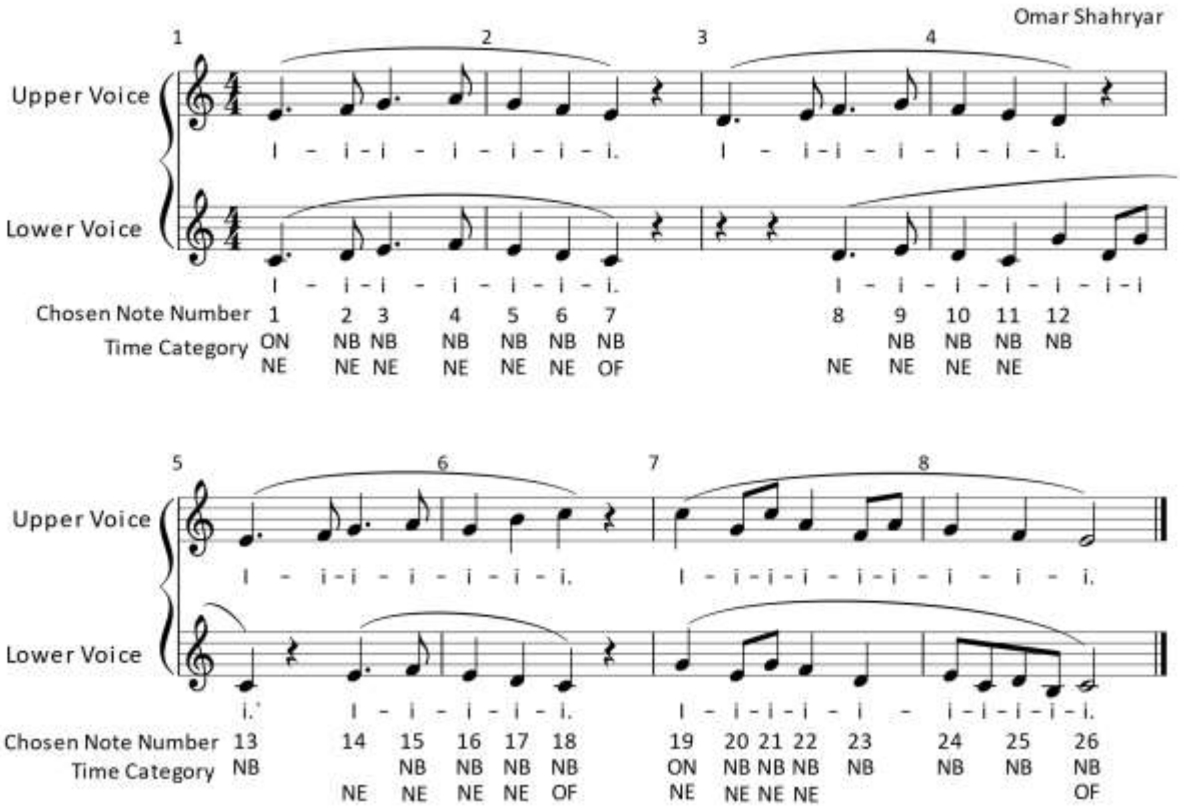


Figure 3.5 *Duo exercise composed for the pilot study presented in chapter 4 and used for the evaluation of the algorithm tested in this chapter as well as the major study described in chapter 5. The figure shows the notes chosen (numbered in ascending order) for the analysis of synchronization, reported in chapters 4 and 5, alongside the four sets of time categories upon which the analysis is based (i.e., onset, ON; note beginning, NB; note ending, NE; and, offset, OF). All notes were used for the evaluation of the reliability of TIMEX in this chapter. The relevant time categories of each chosen note, as shown in the figure, were extracted for both the upper and lower voice.*

3.3.3 Test 2: Reconstruction Process Effectiveness

To assess the precision of the reconstruction process, a set of 40 performances, comprising 40 Lx segments and corresponding audio segments, was randomly selected from the subset of recordings presenting no interruptions in the Lx signals. The group of recordings selected for this test comprised 20 recordings of the upper voice and 20 of the lower voice of the piece, with an equal number of performances for each duo**.** Each clip averaged 30 ms in length and the total length of the performances selected was around 20 minutes. The evaluation of the reconstruction process was done by comparing the reconstructed Lx signal with the corresponding raw Lx signal. The raw Lx values after the first note and before the last note performed were initially deleted from the Lx recordings. Then, the reconstruction process was run based on the audio signal and raw Lx values remaining after deleting the values associated with the first and last note. Eventually, the raw values were compared with the reconstructed recordings. Results show an average margin of error of 0.034 %; the margin of error (E) was first computed for each data point as follows,

$$E = \frac{|V_{\mathrm{raw}} - \mathrm{V_{rec}}|}{\mathrm{V_{raw}}} \tag{9}$$

and then averaged across the entire sample. V_{raw} represents the raw value extracted from the Lx signal, whilst the V_{rec} is the value reconstructed from the algorithm based on the shape of the audio signal.

3.3.4 Test 3: Algorithm Performance

The effectiveness of the algorithm was tested on a subset of 28 Lx recordings, including seven recordings for each voice (i.e., upper and lower voice) and duo (i.e., duo 1 and duo 2). These recordings were randomly selected from the entire pool of data collected for each voice/duo. This test relied solely on the Lx signal and the audio signal was not considered, since the data extraction is mostly based on the Lx recording, and the audio recording is only scrutinised in the very few cases of Lx discontinuous signal, as reported in section *3.3.2*. The data selected for the evaluation of the algorithm performance includes 728 NBs, 728 NEs, 112 ONs, and 112 OFs, with a total of 1680 timing extractions. Each audio file was approximately 25 seconds long, and the total length of the audio clips was about 12 minutes, which is much longer than the about 3 minutes of singing recordings used in the Music Information Retrieval Evaluation eXchange (MIREX) onset detection task, which represents a very common comparison in this field of research to assess the performance of a novel algorithm for onset detection (Friberg, Schoonderwaldt, and Juslin 2007; Dixon 2006; Chang and Lee 2014; Zhou and Reiss 2007).

Recordings selected for the tests in the current study were manually cross-annotated by three experts in the analysis of the singing voice. These experts were third-year PhD students in Music Technology at the Department of Electronic Engineering of the University of York. They all had extensive experience in acoustic analysis including voice analysis using the Praat software. They were external to this investigation, and marked the beginning and ending of each note using Praat software (Boersma 2001; Boersma and Weenink 2013). Experts used the same software setup as each other, displaying a spectrogram and a waveform with a fixed time window, and a tier for hand annotations; this display setup also gave the experts the chance to listen to the recordings. Markings were applied to recordings of the two-part performances sampled at 48 kHz and post-processed with a time step of 1 ms. This time step setting was chosen to allow the detection of small asynchronies in the order of tens of milliseconds, such as those found in the literature of music ensemble performances.

The evaluation procedure followed that described in MIREX 2016 for onset detection. A tolerance value was set to ± 50 ms and the detected times were compared with ground-truth values manually detected by the experts. This is a standard procedure for the evaluation of onset detection algorithms, although the comparison of values detected by the algorithm with those manually detected by experts, and commonly referred to as "ground-truth" values, remains ambiguous and subjective as there can be no true objective value. A large time displacement of 50 ms is a well-known criterion in the field of onset detection that takes into account inaccuracy of the hand labelling process (Bello et al. 2005). In addition, a small-time window of 10 ms was also chosen to detect small asynchronies in the synchronization during professional ensemble performances. The mean of the standard deviations for the manual annotations computed across the three experts was 59 ms.

For a given ground-truth onset time, any extracted value falling within the tolerance time window of 10 or 50 ms was considered correct detection (CD). If the algorithm detected no value within the time window, the detection of that ground-truth time was reported as a false negative (FN). Detections outside all the tolerance windows were counted as false positives (FP). The performance of the detection method was evaluated based on the three measures commonly used in the field of onset detection: Precision (Pr), Recall (Re), and F-measure (F). The Precision measures the probability that the detected value is a true value, thus calculating how much noise the algorithm provides. The Recall indicates the probability that a true value is identified, therefore measuring how much of the ground truth the algorithm identifies. The F-measure represents the overall performance, calculating the harmonic mean of Precision and Recall. The measures are computed as follows:

$$Pr = \frac{N_{\mathrm{cd}}}{\mathrm{N_{cd}+N_{fp}}} \tag{6}$$

$$Re = \frac{N_{\mathrm{cd}}}{\mathrm{N_{cd}+N_{fn}}} \tag{7}$$

$$F = \frac{2\mathrm{PrRe}}{\mathrm{Pr}+\mathrm{Re}} \tag{8}$$

N_{cd} is the number of correct values detected by the algorithm; N_{fp} the number of false values detected; N_{fn} the number of missed values. Because files were cross-annotated by three experts, the mean Precision and Recall rates were defined by averaging Precision and Recall rates computed for each annotation. The overall results are reported in **Table 3.2**.

Table 3.2 *Performance of TIMEX*

Tolerance	Precision	Recall	F-measure
50 ms	65 %	97 %	78 %
10 ms	23 %	89 %	36 %

TIMEX achieved higher results in all measures than the best-performing algorithms for the singing voice from MIREX (2016) with the same threshold of 50 ms, although based on a different data set and extracting different timing categories, such as onsets in MIREX and onsets/offsets/beginnings/endings by TIMEX.

The full data set of detection errors was scrutinized to investigate how false positive and false negative errors were distributed across performers and over the duration of the pieces. As shown in **Table 3.3**, the detection errors, computed with a tolerance level set at 10 ms, varied across the four performers: the total number of false negatives found for singer 2 performing the upper voice was approximately half that of singer 1 performing the same piece, and the total number of false positives for singer 3 performing the lower voice was less than those found for singer 4 performing the lower part. These results suggest that singers might have a particular singing technique that affects the performance of the algorithm.

Table 3.3 *Distribution of detection errors across performers. False negatives and false positives were averaged across performances. Tolerance level set at 10 ms.*

Performer's Part	False Negatives	False Positives
S1 Upper Voice	5.4	136.1

S2 Upper Voice	2.2	130.3
S3 Lower Voice	6.3	119.6
S4 Lower Voice	4.0	158.8

As shown in **Figure 3.6**, the total number of false positives was distributed similarly across the course of the piece. However, false negatives were more likely to occur when the note being analysed was a semitone from the previous note (as found regarding notes 1-2, 6-7, 20-21 and 29-30 of the upper voice, and 4-5, 16-18, 29-30 of the lower voice) or for intervals greater than a 3rd (as found regarding notes 22-24 of the upper voice, and 11-14 and 26-27 of the lower voice).

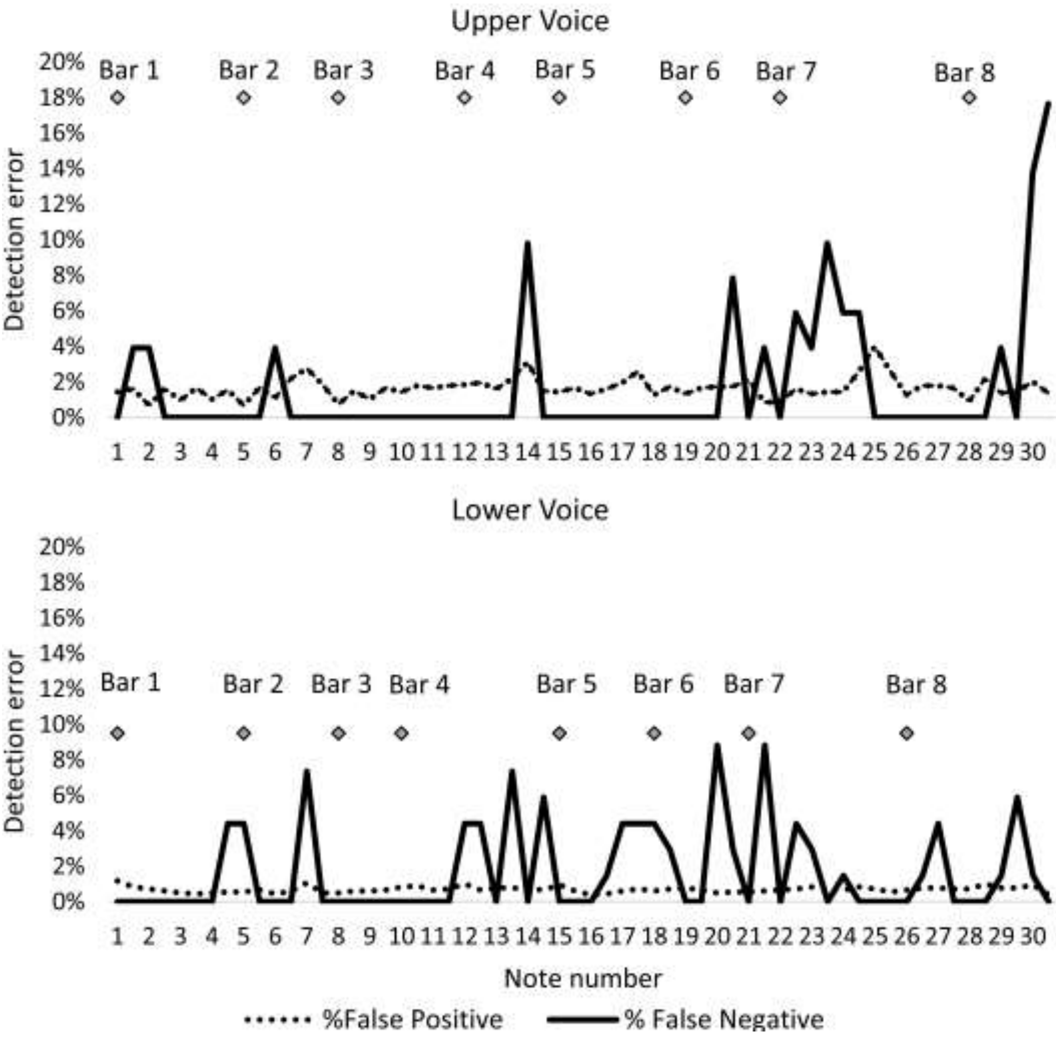


Figure 3.6 *Distribution of percentage detection errors computed at the beginning and ending of each note across the course of the piece.*

3.4 Discussion

The aim of this chapter was to present and test a new algorithm, termed TIMEX, that extracts onsets and offsets of phonation as well as note beginnings and note endings within a legato phase from monophonic recordings of ensemble singing, based on the combined application of Lx and acoustic analysis to track the individual contribution of each singer.

The use of Lx allowed the identification of the contribution of individual voices, avoiding the complication of polyphonic recordings. The audio recordings were used as supporting data in the rare cases in which the Lx signal was too weak to be processed. This set-up was very successful: the Lx signal failed on only 0.7 % of the entire set of recordings, during which the analysis had to rely on the acoustic signal, which could potentially suffer from audio bleed from the other singers. In order to ensure accurate and reliable recordings of vocal fold vibration in the Lx signal, the proper placement of the electrodes is fundamental. The electrodes should be placed in the thyroid region behind the vocal folds in the middle of each thyroid lamina (Colton and Conture 1990). Furthermore, consideration should be given to the fact that the Lx signal may be too weak or noisy to be reliable for use on certain populations, including children (Howard 2009), sopranos (Pabst and Sundberg 1993), and when a thick layer of subcutaneous tissue is present in the neck (Askenfelt, Gauffin, and Sundberg 1980; Colton and Conture 1990)

The evaluation of TIMEX in the present study showed an overall F-measure of 78 % within a tolerance window of 50 ms. Whilst the issues of singing onset detection cannot be considered solved by this system, its potential is promising in light of the *state-of-the-art* techniques presented at MIREX in 2016 yielding F-measures of around 60 %. Direct comparisons with other methods cannot be made unless the same data set is used; comparative evaluations are planned in the future.

The algorithm presented in this paper is based on the f_o profile extracted from Lx and acoustic recordings. It has been developed on the basis of a purely mathematical definition of a local max/min, with the addition of a series of rules to ignore points that the definition would retain but would not represent a change of note in the score being performed. The rules have been conceived based on the issues encountered during the first processing attempts, such as local spikes, vibrato, Lx signal interruptions, and onset/offset fluctuation range. Each of these rules is associated with the definition of a threshold parameter to enforce the rule, which has been tweaked by trial and error to provide the most accurate results, comparing the output of the algorithm for the selected recording to the score that was performed. This algorithm allows mostly for consistency and repeatability of the results, though there was trial and error tweaking, and addresses much of the subjectivity and laboriousness typical of the manual detection.

When testing the algorithm, the same parameters were used for the four semi-professional singers performing the piece, and for the upper and lower voice parts. The fluctuation threshold and the vibrato frequency threshold can be expected to be different for opera singers, who might exhibit a larger vibrato extent than the singers who participated in the study. Optimal values regarding rest, fluctuation and spiking range are expected to vary across pieces, especially if the tempo and duration of rests and notes at the beginning and ends of phrases (and therefore onset and offset of phonation) are very different from the two-part piece used for this set of recordings.

Other avenues of research should consider issues relating to the small fluctuations within the onsets; TIMEX limits the detection to local max/min points, whilst the ground truth considers also the steepness of the f_0 profile, by detecting onsets based on the rate of change of the curve. This could be addressed by developing the algorithm further using the second derivative of the waveform in addition to the max/min points. A future direction of this research should also consider the analysis of the singing voice with lyrics. Moreover, it is reasonable to expect that this algorithm will work well with percussive instruments, although they would probably require the use of different thresholds for the same rules.

Finally, TIMEX runs on Microsoft Excel and relies on two sets of data consisting of f_o estimates and related time stamps of the Lx and audio signals; these two data sets currently need to be extracted from the recordings using different software, such as Praat (Boersma 2001; Boersma and Weenink 2013), and then manually imported in Excel as tabular lists of data. This process can be improved and manual processes avoided, re-writing TIMEX using a more powerful programming language, such as MATLAB, allowing for the automatic extraction of f_0 and time stamps data lists, and faster and more stable data processing.

3.5 Conclusions

This chapter has described and tested a novel algorithm, termed TIMEX, that automatically detects onsets and offsets of phonation as well as note beginnings and endings within a legato phrase from singing recordings, based on the combined application of Lx and acoustic analysis. The combined application of these two techniques represented a successful set-up to track the individual contributions of singers in duo recordings. The evaluation of the new algorithm, applied to a set of singing duo recordings, has shown an overall performance of 78 % within a tolerance window of 50 ms compared with manual annotations performed by three experts. Results seem promising in light of the *state-of-the-art* techniques presented at MIREX in 2016 yielding an overall performance of around 60 %. The algorithm allowed for some consistency and repeatability of the extraction of timing information, solving many of the issues regarding the subjectivity and laboriousness of the manual detection of onsets. In the next chapter, this algorithm is applied in a pilot study testing a protocol (i.e., algorithm implementation, stimulus, procedure, and analytical framework) for the study of the roles of visual contact and leadership in the temporal coordination between singers during singing duo performances.

4 A Pilot Study of the Impact of Visual Contact and Leadership Instruction on Ensemble Singing

4.1 Introduction

The previous chapter presented and tested a novel algorithm for the automatic detection of note timings in singing recordings that proved to be a successful and valuable tool for the extraction of onsets/offsets of phonation and note beginnings/endings within a legato phrase, from polyphonic singing recordings. In the current chapter[5], this algorithm is implemented in a pilot study evaluating a protocol for the analysis of the impact of visual contact and the instruction to act as leader or follower on the synchronization between singers in the context of singing duos, since the protocols used in previous investigations were mostly applied within instrumental ensembles.

Current quantitative investigations analysing the impact of visual contact (Keller and Appel 2010; Goebl and Palmer 2009; Kawase 2014; Bishop and Goebl 2015) and leadership instruction (Goebl and Palmer 2009; Zamm, Pfordresher, and Palmer 2015) on synchronization have often provided inconclusive results, as shown in section *2.2.1* (for visual contact) and section *2.2.2* (for leadership instruction), and have focused on instrumental ensemble performances. The systematic study of synchronization in ensemble singing currently lacks investigation, and so does the impact of visual contact and leadership on synchronization in ensemble singing.

The presence/absence of visual contact between singers during singing performances may play an important role in interpersonal synchronization, as a singer's body is their instrument. Whilst some instrumentalists such as piano players might have the additional factor of looking sometimes at the music instrument (which is physically independent from their own body) to boost expressive ensemble performance, a singers' vision might be mostly focused on the score and/or co-performer(s)' and conductor's body gestures during singing performances. The leadership instruction may also be important in the context of ensemble singing, since small ensembles do not have a choral director and therefore there might be a decision made

---

[5] Most of this chapter, including figures and tables, is reported in D'Amario, Daffern, and Bailes (2018b), and is © the authors, licensed CC-BY.

that leadership will be shared among members of the ensembles, as found in observational studies of a professional singing ensemble (Lim 2014) and an 11-men singing ensemble (Page-Shipp, Joseph, and van Niekerk 2018).

This pilot study serves as a test of a protocol for the subsequent experiment, conducted with a larger sample of duos and described in the following chapter, investigating the roles of visual contact and leadership instruction during singing duo performances. In addition, the implementation and testing of a protocol for the analysis of the roles of visual contact and leadership instruction in ensemble singing provides also the foundations for the study of the evolution of synchronization across rehearsals reported in chapter *6*.

Although this study is explorative in nature, it was hypothesised that visual contact does not affect synchronization during the performance of a two-part piece featuring a regular metre, based on previous investigations conducted with piano duos using similar stimuli (Keller and Appel 2010; Goebl and Palmer 2009). There was no specific hypothesis to test in relation to the impact of the leadership instruction, since previous investigations report apparently contrasting results regarding the effect of leadership instruction on temporal coordination between musicians, which might be due to the score used in the study (Goebl and Palmer 2009; Zamm, Pfordresher, and Palmer 2015).

4.2 Method

4.2.1 Participants

Four undergraduate singing students (three females and one male; age $Mdn = 20.3\,yo$, $Range = 1.7\,yo$) were recruited from the Department of Music at the University of York. Singers had at least seven years' experience performing in a singing ensemble ($Mdn = 9\,y; Range = 8\,y$), but they had not sung together prior to the experiment. They reported having normal hearing and not having absolute pitch.

4.2.2 Stimulus

A vocal duo exercise was composed for this study, featuring mostly a homophonic texture to allow investigation of the synchronization per note, as shown in **Figure 3.5**. The upper voice has a range of a 7th, and the lower voice a range of a 5th. The upper voice features a higher tessitura than the lower voice. The piece comprises eight bars, sung legato to the vowel /i/ and separated by rests, thus representing a short, stand-alone piece. The choice of the vowel was left to the first duo taking part to the study, who chose the vowel /i/. For consistency, the same vowel was used with the second duo taking part in the pilot, and also with the remaining singing ensembles involved in this thesis (see the singing quintet in chapter *6*). Lyrics, consonants and repeated notes within the legato phrase were not used, and the presence of semitones was limited to facilitate the analysis of synchronization based on the f_o estimates (for more information about the choices upon which the composition of the piece was based, see section *3.3.1*).

4.2.3 Apparatus

Participants were invited to sing in a recording studio at the University of York, treated with absorptive acoustic material. The room was $4.1\ m \times 3.4\ m$, the ambient noise level was 32 dB(A), and the RT60 reverberation time was 0.15 s. Therefore, the room was an environment not alien to the singers in terms of acoustic and look, but a typical recording studio, relatively dead, with acoustic panelling on the walls and recording equipment. Singers wore head-mounted close proximity microphones (DPA 4065), placed on the cheek at approximately 2.5 cm from the lips, and electrolaryngograph (Lx) electrodes (from Laryngograph Ltd www.laryngograph.com) placed on the neck positioned either side of the thyroid cartilage, as shown in **Figure 4.1.** One stereo condenser microphone (Rode NT4) was placed at equal distance in front of the singers at approximately 1.5 m from the lips. The stereo microphone was a twin cardioid microphone pointing between the singers, with its main axes pointing at the two singers. This stereo microphone was chosen as being most sensitive to the area in front of the microphone capsule, where the singers stood, while picking up minimal and marginal noise from the rear and sides, respectively. The six outputs (two Lx, two head-mounted microphones, one stereo microphone comprising right and left channel) were connected to a multichannel hard disk recorder (Tascam DR680), as shown in **Figure 4.2**, and recorded at a sampling frequency of 48 kHz and 24-bit depth.

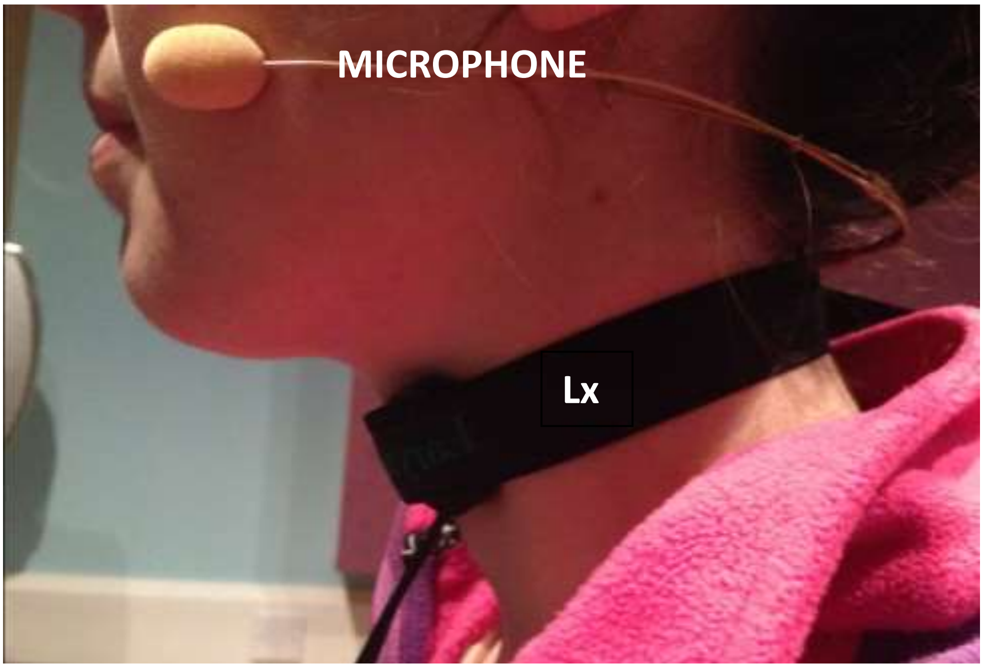


Figure 4.1 *Example of the application of head-mounted microphone and electrodes to extract Lx and audio data from a singer.*

4.2.4 Design

The study consisted of a within subject design in which participants were asked to sing the piece in the following four conditions, applied in a randomised order:

- VC_UpperVoiceL: with visual contact, and upper voice designated leader and lower voice (LowerVoice) follower
- VC_UpperVoiceF: with visual contact, and upper voice designated follower and lower voice leader
- NVC_UpperVoiceL: without visual contact, and upper voice designated leader and lower voice follower
- NVC_ UpperVoiceF: without visual contact, and upper voice designated follower and lower voice leader

Each condition was presented three times, resulting in 12 takes; each take consisted of four repeated performances of the piece, resulting in a 4 (conditions) × 3 (takes, i.e. repeated performances of each condition) × 4 (repeated performances within each take) design featuring a total of 48 repetitions of the piece per duo.

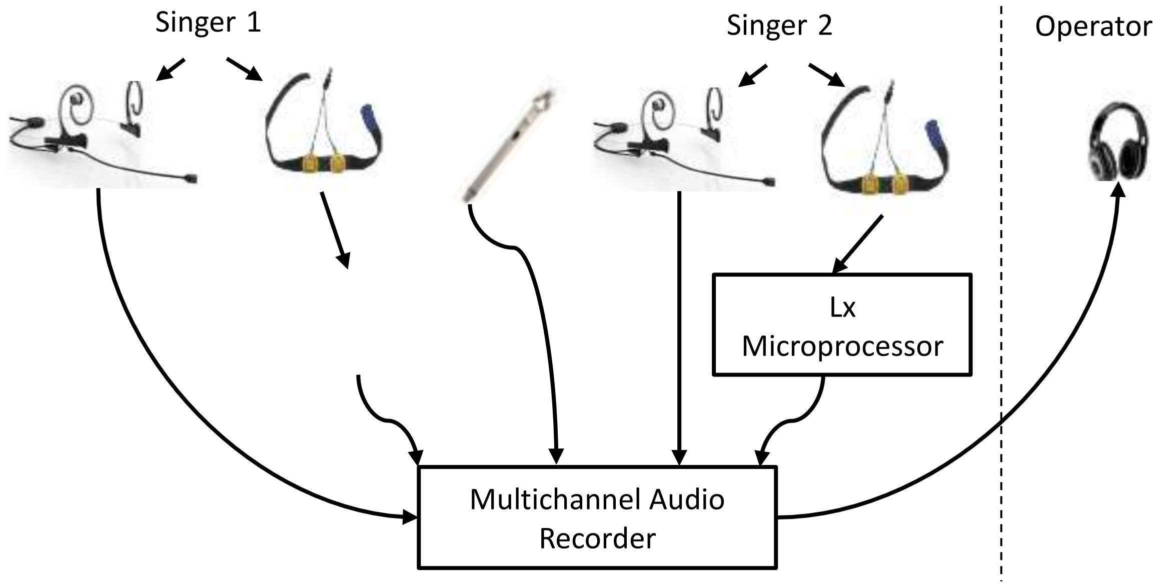


Figure 4.2 *Example of a typical experimental set-up.*

4.2.5 Procedure

Singers received the stimulus material one week prior to the experiment, to practise the piece. On the day of the experiment, firstly participants were asked to fill in a background questionnaire and consent form. Then, head mounted microphones and Lx electrodes were placed on each singer and adjusted. The correct placement of the Lx electrodes was verified by checking the signal on the visual display and listening over headphones. The microphones were adjusted for the sound pressure level of each participant to avoid clipping. Singers were invited to familiarize themselves with the piece for 10 minutes, singing together from the score, while listening for 10 seconds to a metronome set at 100 beats per minute (BPM) before starting to rehearse. If the singers were able to perform the piece without errors, the four conditions and associated 12 takes were then presented; otherwise, they were allowed to practise the piece for 10 more minutes and then the test was repeated. Once the musicians passed the performance test without errors with the score, each singer was assigned the role of leader or follower; these roles were then reversed according to UpperVoiceL and UpperVoiceF conditions. Signs labelled “leader” and “follower” were placed on the floor in front of the participants, to remind them of their roles. Each singer only had one assigned part/musical voice, i.e. a given singer only ever sang the upper or the lower part. Singers were invited to face each other at a distance of 1.5 m in the visual condition and to face away from each other at the same distance in the non-visual contact condition. Singers were not aware of the purpose of the study. The 12 takes were recorded singing by heart with short breaks between each of them. The experiment lasted approximately one hour.

Ethical approval for the study was obtained from the Physical Sciences Ethics Committee (PSEC) at the University of York (UK), with reference D'Amario151127.

4.2.6 Analysis

The analysis of synchronization relied on the Lx and head-mounted microphone recordings, thus the stereo recordings were not scrutinized for the study of temporal coordination, as they do not allow tracking of the individual contributions.

For each recorded performance, two sets of data including the audio waveform from the head-mounted microphones and the Lx waveform were imported into Praat (Boersma 2001; Boersma and Weenink 2013) as .wav files and the fundamental frequency (f_o) values were extracted with a time step of one ms. These data were imported into Microsoft Excel 2016 in the form of a tabular list of data points, including the f_o in Hertz and corresponding timestamp in milliseconds with a time step of 1 ms. Asynchronies were then calculated to measure the phase synchrony between singers for onset (ON) and offset (OF) of phonation, and note beginning (NB) and note ending (NE) within a legato phrase, related to the selected notes, as shown in **Figure 3.5**. Those notes were chosen as being relevant to synchronization. The phase asynchrony was computed subtracting the follower's timestamp values from the leader's (leader minus follower) related to NB, NE, ON and OF of the selected notes. Negative values show that the leader preceded the follower, while positive values indicate that the follower is ahead of the leader.

The detection of ON, NB, NE, and OF was automated through the application of the TIMEX algorithm described in the previous chapter. This event detection method was visually validated for the entire data set by the author (SD). Soft phonation was specifically scrutinized in respect to the Lx signal, which might not pick up very small vocal fold vibrations when the amplitude is very small (Sten Ternström and Johan Sundberg, personal communication, June 2017). In cases whereby the phonation was too soft to be picked up by the Lx signal, the timing detection was mostly based on the acoustic recording, which might potentially suffer from bleed from co-performers. Cases of soft phonation, however, were very scarce. In addition, occasional note errors due to the musicians entering or delaying the notes for more than 50

% of the expected value were also investigated by comparing the f_0 values and the audio recording with the notated score. Takes in which a note error occurred were excluded from the analysis. The error rate was less than 1 %. In addition, during the author's validation, it was noted that the presence of few semitones slightly affected the performance of TIMEX. Although the vibrato frequency parameter described in the previous chapter was supposed to differentiate the oscillations of the vibrato from that of a semitone, this was not always the case. A manual detection in these cases was necessary. The manual changes computed by the author in relation to the semitones were less than 2 %. Manual changes for an additional 34 % of the data set were also necessary in cases in which the singers performed a vibrato (within the same note) with an extent larger than a semitone; in these cases, the algorithm identified the frequency oscillations of the vibrato as note beginnings/endings, failing to detect the vibrato. The overall manual changes related to errors associated with timing, semitones, and vibratos were around 37 %.

Outliers were identified based on the MAD (Median Absolute Deviation), and asynchronies that fell more than 2.5 absolute deviations from the median were excluded. This approach is the most robust method to detect outliers, when the distribution is not normal and outliers are present (Leys et al. 2013), as in this case.

Four aspects of synchronization were measured:

- Precision of interpersonal synchronization, as indexed by the mean of absolute asynchronies, and expressed in milliseconds.
- Consistency of synchronization, as indexed by the standard deviation (SD) and the coefficient of variation (CV) of the absolute asynchronies. The SD of asynchronies is expressed in milliseconds; the CV, defined as the ratio of the standard deviation to the sample mean, is reported as percentage.
- Tendency to precede or lag a co-performer, as indexed by the median (Mdn) of signed asynchronies, and expressed in milliseconds.

The main synchronization characteristics were firstly investigated across the course of the 48 repeated performances. For this purpose, descriptive statistics were calculated for duo, and synchronization measure (i.e., precision, consistency and tendency to precede) across a multiple-step averaging process that comprised: i) pooling together the relevant time category of each note and averaging across each note within the same performance; ii) pooling together each note and averaging across each performance; finally, averaging across performances pooling together the 48 performances. This computation was repeated for each synchronization measure/duo.

Data were also investigated for evidence of changes in interpersonal synchrony across the course of the 48 repeated performances, that might have been induced by fatigue or learning. For this purpose, descriptive statistics were calculated for each performance, synchronization measure (i.e., precision, consistency and tendency to precede) and duo, pooling together the relevant time category of each note, and then averaging across each note. This computation was repeated for each performance/synchronization measure/duo, and provided 48 asynchrony values (i.e., one per performance) for each synchronization measure and duo.

The impact of the presence/absence of visual contact between singers and the leadership instruction was evaluated through descriptive analyses and inferential paired tests including dependent paired *t*-tests and Wilcoxon signed rank tests. *T*-tests were chosen to analyse differences between means within the absolute asynchronies data sample, whilst Wilcoxon tests were selected to assess median differences across signed asynchronies. These statistical tests were run for each time category (i.e., ON, NB, NE, and OF), synchronization measure, condition, and duo, pooling together the asynchronies of each note, and then averaging across each performance. This process provided 48 asynchronies values computed for each time category, condition (i.e., 24 VC – 24 NVC, and 24 UpperVoiceL – 24 UpperVoiceF), synchronization measure and duo, which were then entered in SPSS (IBM SPSS Statistics v. 24) for the relevant statistical tests.

4.3 Results

4.3.1 Synchronization Characteristics

The analysis of the overall synchronization was computed regardless of performance condition for each duo across the 48 repeated performances, and results reported in **Table 4.1**. One sample t-tests for difference than 0 conducted on the median signed asynchronies, to evaluate the significance of the tendency to precede/lag, demonstrate that the designated leader of duo 2 was significantly ahead of the designated follower $[Mdn = -10\ ms, t(47) = -2.6, p = 0.005, r = 0.52]$.

Table 4.1 *Summary of the mean and median values computed for each duo (i.e., duo 1 and duo 2) and synchronization parameter (i.e., precision quantified by mean absolute asynchronies, consistency measured by SD and CV of absolute asynchronies, and tendency to precede/lag quantified by median signed asynchronies). In square brackets [], SD values of precision and consistency values alongside the interquartile range (IQR) values of median signed values are reported. CV values are expressed as percentage of the related sample mean, whilst the other values are expressed in ms. The levels of p-values for the significant tendency to precede/lag resulting from one sample tests for difference than 0 are also reported* ($* p < 0.05$).

	Precision (M)	**Consistency (SD)**	**Consistency (CV)**	**Tendency to precede/lag (Mdn signed)**
Duo 1	61 [12]	41 [9]	70 [9]	-7 [30]
Duo 2	67 [12]	41 [8]	60 [9]	-10* [37]

In addition, the analysis of the asynchronies across the course of the 48 repeated performances showed that there were no discernible changes in interpersonal synchronization during the course of the 48 performances for duo 1 or duo 2, as shown in **Figure 4.3** and **Figure 4.4**.

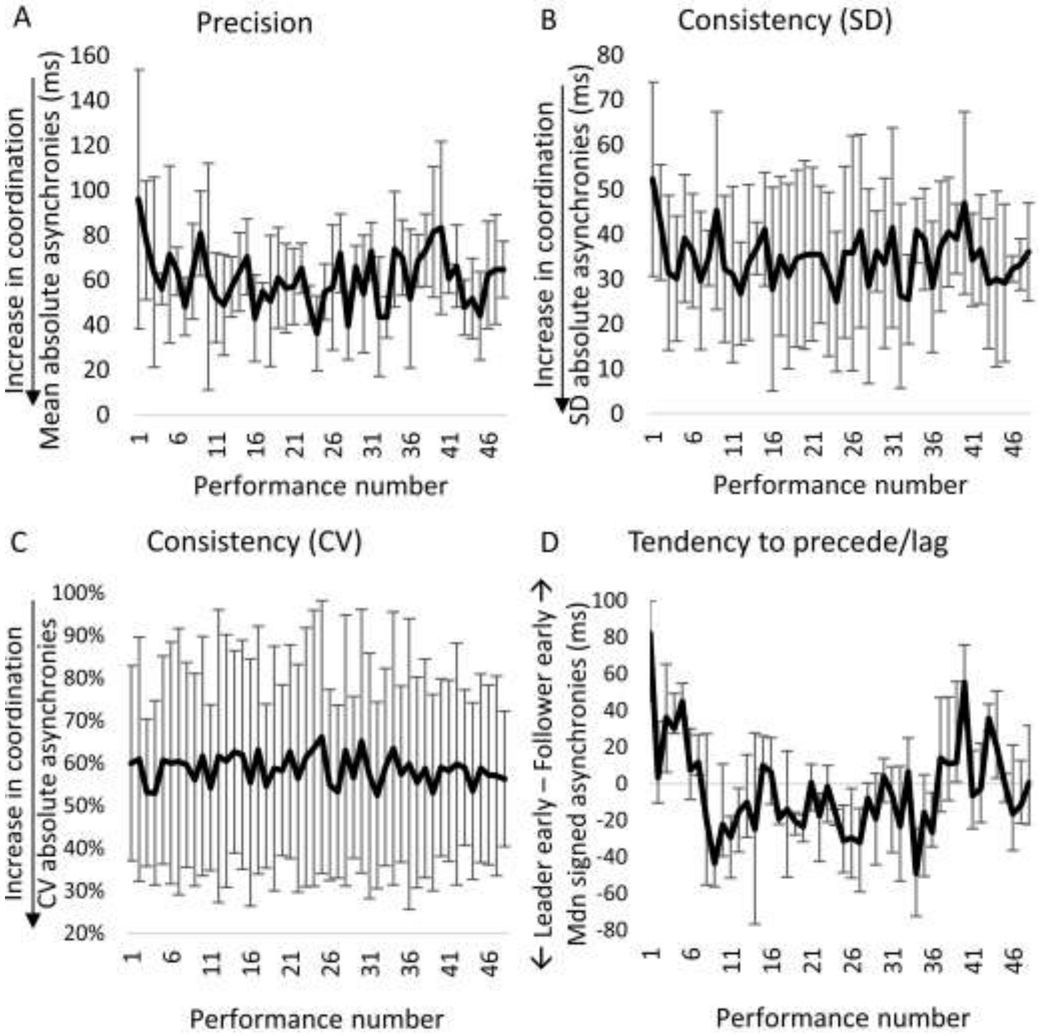


Figure 4.3 *Measures of interpersonal synchrony for duo 1 as precision, consistency and tendency to precede per performance. Error bars of precision and consistency represent standard error of the mean, whilst error bars of the tendency to precede indicate interquartile range of the median.*

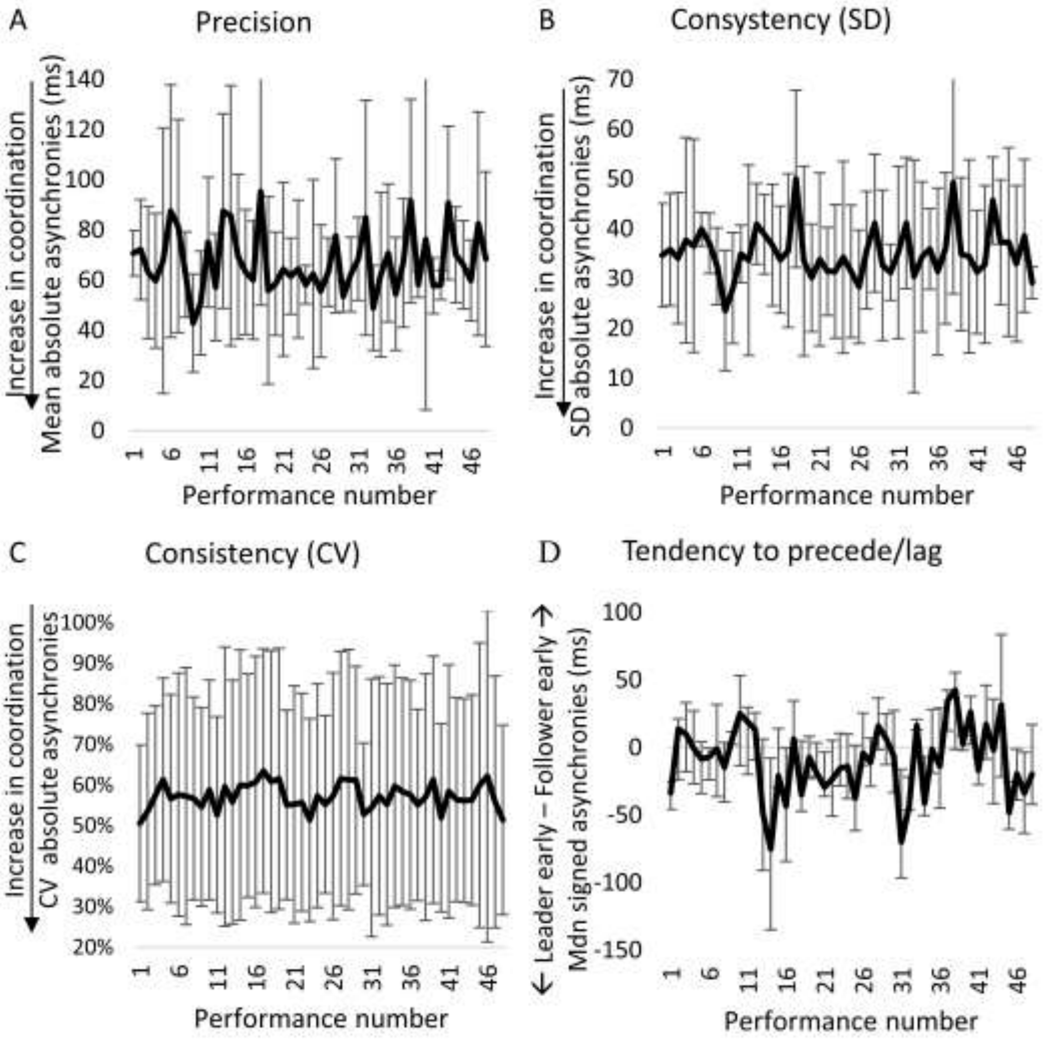


Figure 4.4 *Measures of interpersonal synchrony for duo 2 as precision, consistency and tendency to precede per performance. Error bars of precision and consistency represent standard error of the mean, whilst error bars of the tendency to precede indicate interquartile range of the median.*

4.3.2 Visual Contact

Results, using Bonferroni correction for multiple comparisons, of the impact of visual contact on synchronization are summarized in **Table 4.2**.

4.3.2.1 Duo 1

Mean, SD and CV of absolute asynchronies and median of signed asynchronies for duo 1, calculated for ON, NB, NE, and OF during visual contact (VC) and non-visual contact (NVC), are shown in **Figure 4.5**. Results from the paired sample tests showed a significant effect of the presence of visual contact on the NB standard deviation asynchronies $[t(23) = 2.43, p = 0.023, r = 0.45]$. As can be seen in **Figure 4.5B**, synchronization was found to be significantly more consistent in the NVC condition for NB standard deviation asynchronies, compared with the VC condition. No significant effect was found for the remaining paired sample tests conducted across duo 1.

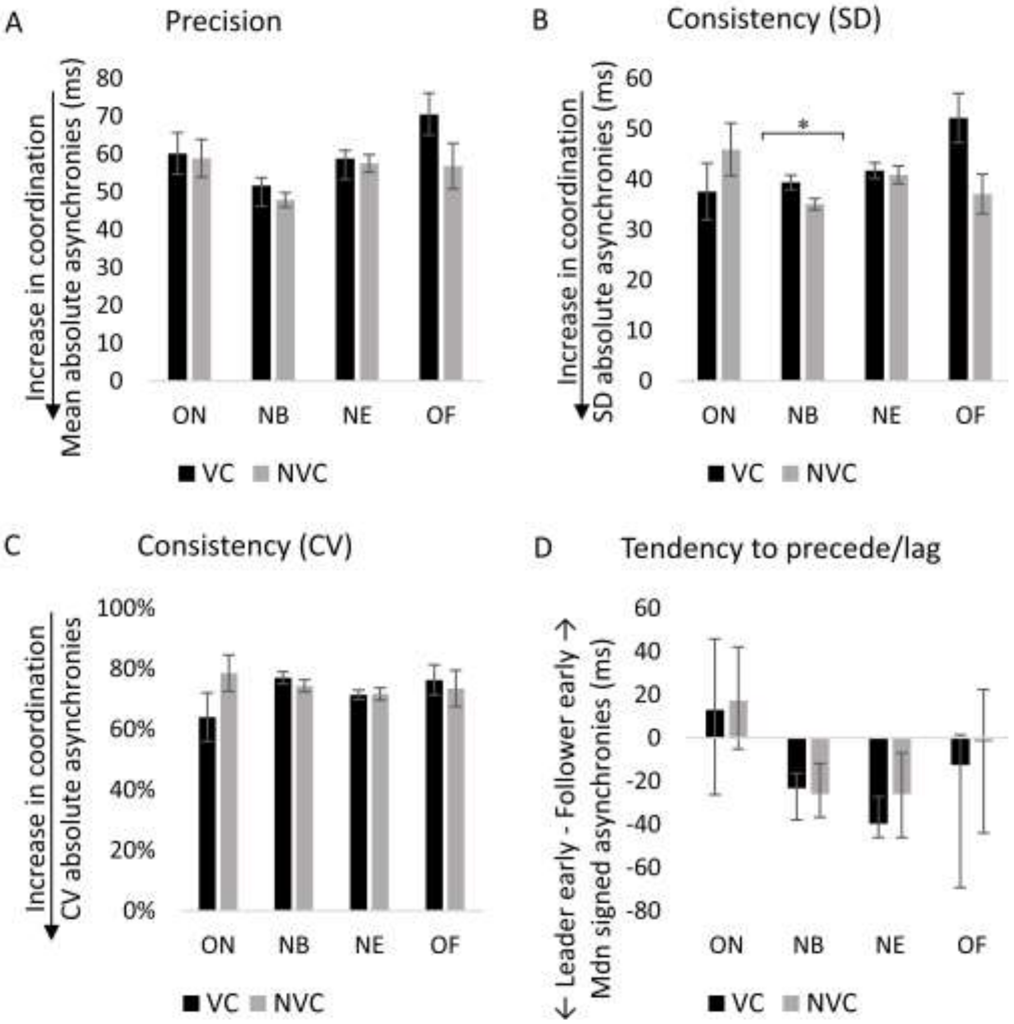


Figure 4.5 *Interpersonal synchronization of duo 1 with visual contact (VC) and without visual contact (NVC) between singers, as indexed by the mean (panel A), standard deviation (panel B), coefficient of variation (CV) of absolute asynchronies (panel C), and median of signed asynchronies (panel D) calculated across ON, NB, NE, and OF. Error bars represent the standard error of the mean for precision and consistency, and the interquartile range for tendency to precede. Smallest values in the precision and consistency of asynchronies indicate an increase in coordination, whilst negative values in the tendency to precede mean that the designated leader is ahead of the follower.* $* p < 0.05$.

4.3.2.2 Duo 2

Median signed asynchronies, and mean, SD and CV of absolute asynchronies for duo 2 are shown in **Figure 4.6**. Paired sample tests were run as for duo 1. The t-test on the mean NB asynchronies highlighted a significant effect of the presence/absence of visual contact $[t(23) = 2.86, p = 0.018,5 = 0.51]$, showing that precision was better in NVC. No significant effect was found for the remaining paired sample tests.

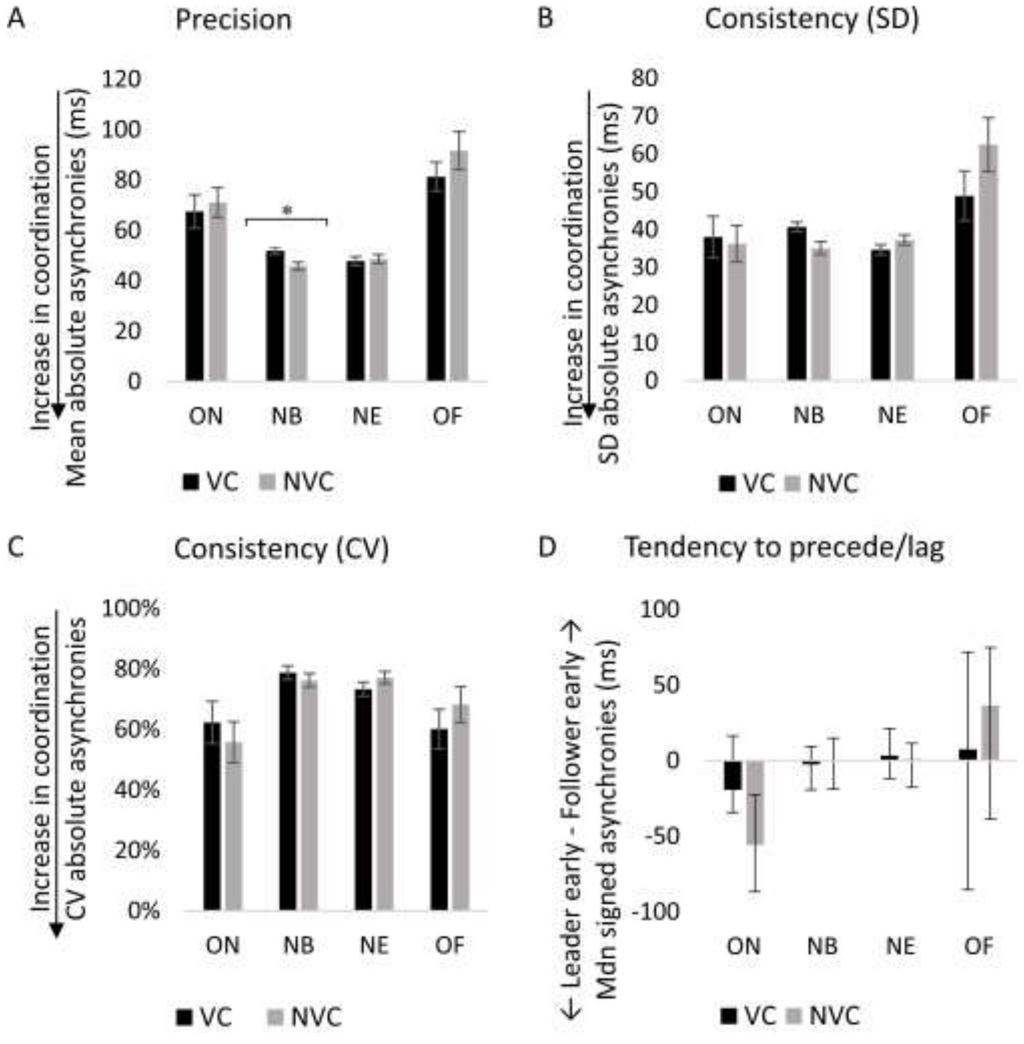


Figure 4.6 *Interpersonal synchronization of duo 2 with visual contact (VC) and without visual contact (NVC) between singers, as indexed by the mean (panel A), standard deviation (panel B), coefficient of variation (CV) of absolute asynchronies (panel C), and median of signed asynchronies (panel D) calculated across ON, NB, NE, and OF. Error bars represent the standard error of the mean for precision and consistency, and the interquartile range for tendency to precede. Smallest values in the precision and consistency of asynchronies indicate an increase in coordination, whilst negative values in the tendency to precede mean that the designated leader is ahead of the follower.* $* p < 0.05$.

4.3.3 Leadership Instruction

Results, using Bonferroni correction for multiple comparisons, of the impact of leadership instruction on synchronization are summarized in **Table 4.2**.

4.3.3.1 Duo 1

Mean, SD and CV of absolute asynchronies, and median signed asynchronies, averaged across the 48 performances in the UpperVoiceL and UpperVoiceF conditions for duo 1, are shown in **Figure 4.7**. Paired sample *t*-tests revealed a significant effect of the instruction to act as leader or follower on both measures of consistency for NB: SD asynchronies $[t(23) = 2.48, p = 0.0021, r = 0.46]$, and CV asynchronies $[t(23) = 2.60, p = 0.016, r = 0.48]$. Consistency of NB synchronization was significantly better when the upper voice was instructed to follow, rather than to lead, as shown in **Figure 4.7B** and **Figure 4.7C**.

Wilcoxon tests revealed a main significant effect of leader-follower instruction on the degree of preceding ON median asynchronies $(T = 60, p = 0.010)$, and NB median asynchronies $(T = 71, p = 0.024)$. One sample *t*-tests conducted on ON for each condition showed that ON median asynchronies when the upper voice was instructed to follow were significantly different from 0 $[Mdn = 36\, ms, t(23) = 3.208, p = 0.004, r = 0.56]$, demonstrating that the designated leader (i.e., the lower voice) was lagging, as indicated by the positive median value (36 ms). One sample *t*-test conducted on NB when the upper voice was instructed to lead showed that NB median values were significantly different from 0 $[Mdn = -20\, ms, t(23) = -6.287, p = 0.000002, r = 0.80]$, indicating that the designated leader (i.e. the upper voice) was ahead of the lower voice, as per the negative median value (-20 ms). Finally, one sample *t*-test conducted on NB asynchronies when the upper voice was instructed to follow showed that NB median data were different from 0 $[Mdn = -27\, ms, t(23) = -11.578, p = 4.5028\mathrm{E} - 11, \mathrm{r} = 0.92]$, demonstrating that the designated leader (i.e., the lower voice) was ahead of the follower, as per the negative median value (-27 ms). In other words, these results demonstrate that when either voice was instructed to lead, the designated leader significantly tended to precede the designated follower at NB. However, when the upper voice was instructed to follow, the designated follower (i.e. the upper voice voice) significantly tended to precede at ON, suggesting that the designated follower was leading against the researcher's instruction.

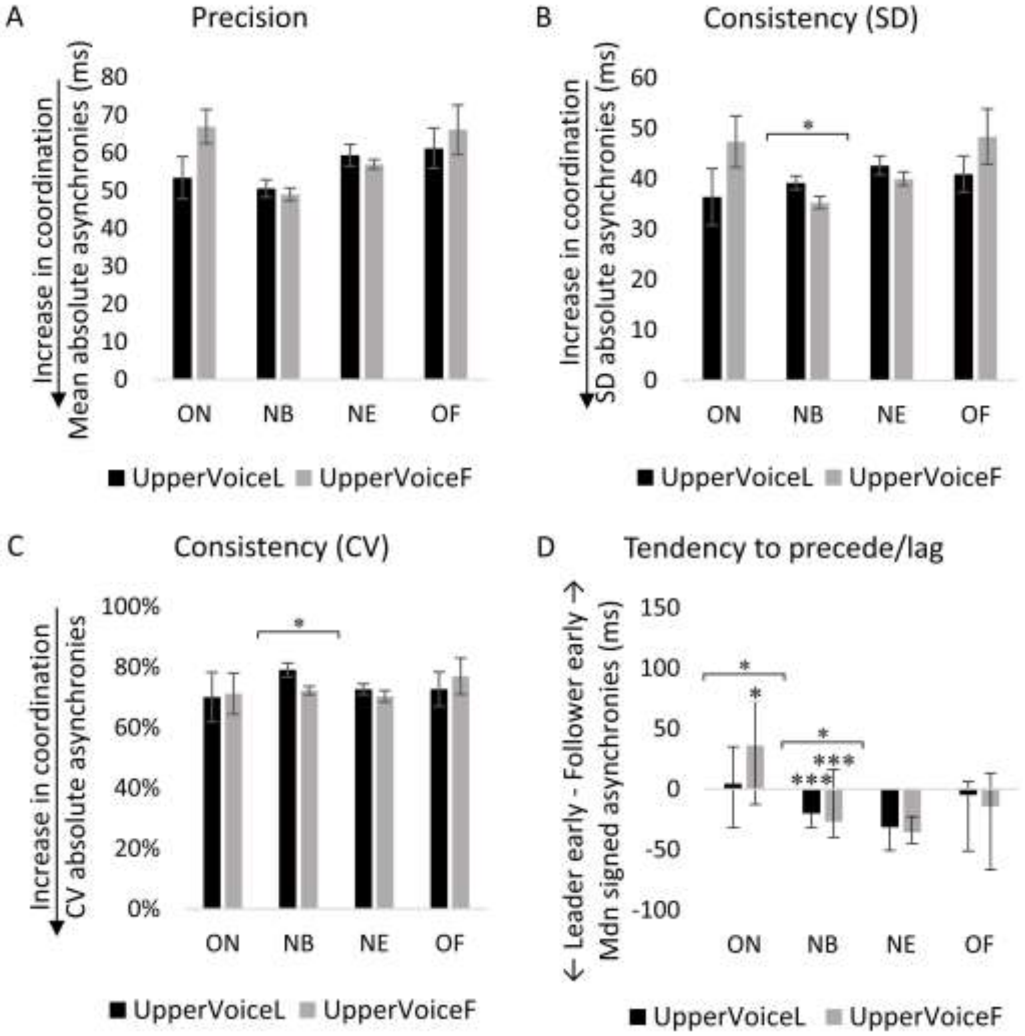


Figure 4.7 *Interpersonal synchronization for duo 1 with the upper voice as the designated leader (UpperVoiceL) or follower (UpperVoiceF), as indexed by the mean (panel A), standard deviation (panel B), coefficient of variation (CV) of absolute asynchronies (panel C), and median of signed asynchronies (panel D) calculated across ON, NB, NE, and OF. Panel D also shows the levels of p-values for the significant tendencies of each voice to precede or lag a co-performer. Error bars indicate the standard error of the mean for precision and consistency, and the interquartile range for tendency to precede. Smallest values in the precision and consistency of asynchronies indicate an increase in coordination,*

whilst negative values in the tendency to precede mean that the designated leader is ahead of the follower. $* \, p < 0.05$; $*** \, p < 0.001$.

4.3.3.2 Duo 2

Median signed asynchronies, and mean, SD and CV of absolute asynchronies computed for duo 2 in UpperVoiceL and UpperVoiceF conditions are shown in **Figure 4.8**. Paired sample tests were calculated as for duo 1. A significant effect of the leader-follower instruction was found on the consistency of NB synchronization, as indexed by: i) SD asynchronies $[t(23) = -4.40, p = 0.0002, r = 0.8]$; and, ii) CV asynchronies $[t(23) = 2.65, p = 0.014, r = 0.48]$. Consistency of NB synchronization was better when the upper voice was instructed to lead and the lower voice to follow.

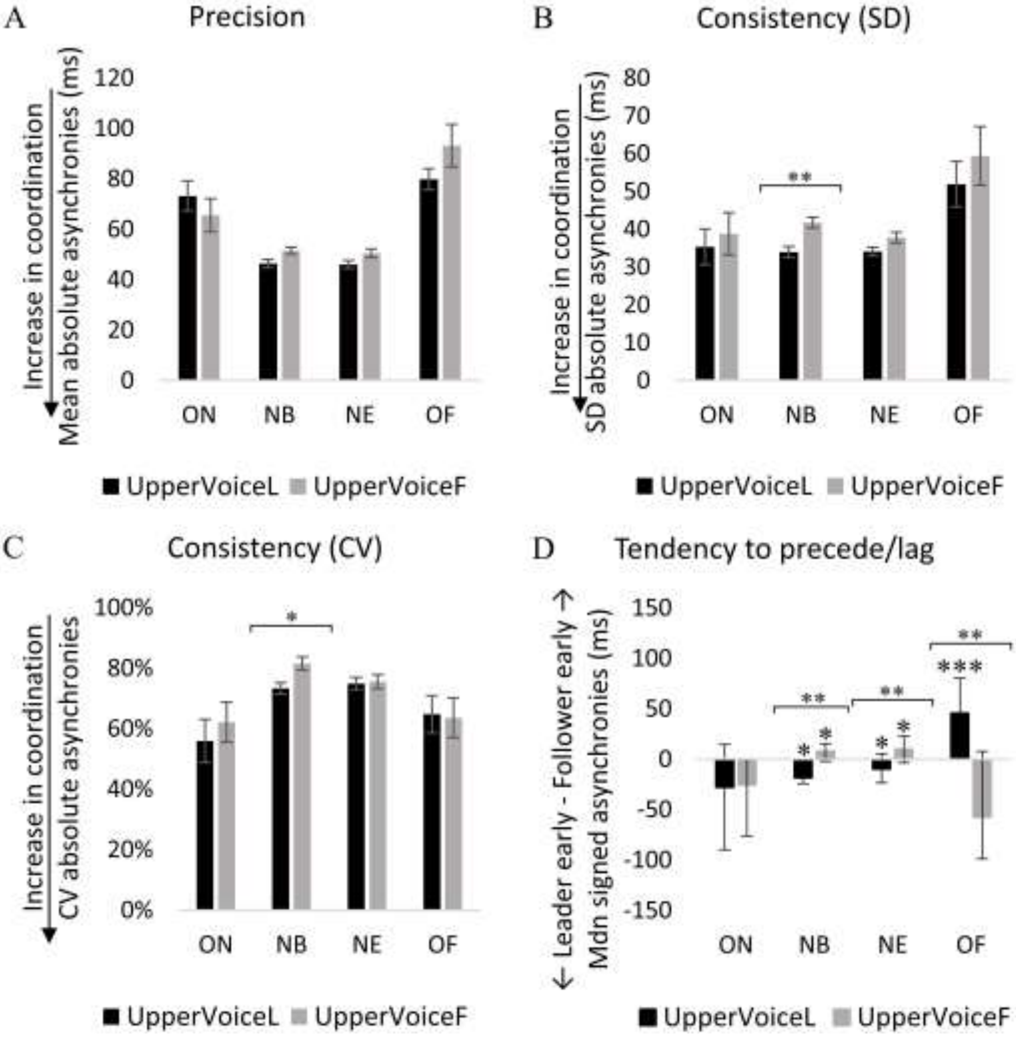


Figure 4.8 *Interpersonal synchronization for duo 2 with the upper voice as the designated leader (UpperVoiceL) or follower (UpperVoiceF), as indexed by the mean (panel A), standard deviation (panel B), coefficient of variation (CV) of absolute asynchronies (panel C), and median of signed asynchronies*

(panel D) calculated across ON, NB, NE, and OF. Panel D also shows the levels of p-values for the significant tendencies of each voice to precede or lag a co-performer. Error bars indicate the standard error of the mean for precision and consistency, and the interquartile range for tendency to precede. Smallest values in the precision and consistency of asynchronies indicate an increase in coordination, whilst negative values in the tendency to precede mean that the designated leader is ahead of the follower. $* p < 0.05$; $** p < 0.01$; $*** p < 0.001$.

Finally, as shown in **Figure 4.8D**, Wilcoxon tests revealed a significant effect of leader-follower instruction on the degree of preceding/lagging: i) median NB asynchronies $[T = 38.5, p = 0.001]$; ii) median NE asynchronies $[T = 33, p = 0.001]$; and, iii) median OF asynchronies $[T = 42, p = 0.002]$. One sample t-tests on median ON, NB, NE and OF were conducted as for duo 1 to observe whether the tendency to precede/lag was significant in each condition. Results showed that NB asynchronies were significantly different from 0 when the upper voice was instructed to lead $[Mdn = -19\, ms, t(23) = -3.564, p = 0.002, r = 0.60]$, and to follow $[Mdn = 8\, ms, t(23) = -2.718, p = 0.012, r = 0.49]$, demonstrating that in both conditions the upper voice tended to precede the lower voice at NB. Results also showed that NE values were significantly different from 0 when the upper voice was instructed to lead $[Mdn = -10\, ms, t(23) = -2.845, p = 0.009, r = 0.51]$, and also to follow $[Mdn = 10\, ms, t(23) = -3.144, p = 0.005, r = 0.55]$, demonstrating that, as for the NB, the upper voice tended to precede the lower voice at NE in both leadership conditions. Finally, OF asynchronies were significantly different from 0 when the upper voice was instructed to lead $[Mdn = 45, t(23) = 4.695, p = 0.00009, r = 0.70]$, demonstrating that the designated leader (i.e., the upper voice) tended to lag the lower voice at OF. In other words, these results demonstrate that when either voice was instructed to lead, the upper voice significantly tended to precede the lower voice at NB and NE. However, when the upper voice was instructed to lead, the designated leader tended to lag at OF. Overall, these results show a complex pattern of leader and follower relationships, rather than a clear separation of roles, which seems to be independent of the researcher's instruction to lead or follow.

Table 4.2 *Summary of the mean and median value per condition showing the levels of p-values for the significant effects resulting from the paired t-tests and Wilcoxon tests* ($\ast\, p < 0.05$; $\ast\ast\, p < 0.01$). *CV values are expressed as percentage of the related sample mean, whilst the other values are expressed in ms.*

	Duo 1		**Duo 2**		**Duo 1**		**Duo 2**	
	VC	NVC	VC	NVC	UpperVoiceL	UpperVoiceF	UpperVoiceL	UpperVoiceF
ON								
Precision (M)	60	59	68	71	53	67	63	66
Consistency (SD)	38	46	38	36	36	47	35	39
Consistency (CV)	60	80	60	60	70	70	60	60
Tendency to lead (Median signed)	13	17	-19	-56	4	36*	-28	-26
NB								
Precision (M)	52	48	52	46**	51	49	46	51
Consistency (SD)	40	35*	41	35	39	35*	34	42**
Consistency (CV)	80	70	80	80	80	70*	70	80*
Tendency to lead (Median signed)	-23	-26	-3	0	-20	-27*	-19	8**
NE								
Precision (M)	59	58	48	49	59	57	46	50
Consistency (SD)	42	41	35	37	43	40	34	38
Consistency (CV)	70	70	70	80	70	70	70	70
Tendency to lead (Median signed)	-40	-26	3	1	-32	-36	-10**	10**
OF								
Precision (M)	70	57	81	92	61	66	80	93
Consistency (SD)	52	37	49	62	41	48	52	59
Consistency (CV)	80	70	60	70	70	80	60	60
Tendency to lead (Median signed)	-12.5	-2	8	36	-4.8	-15	45	-58**

4.4 Discussion

This chapter describes a pilot study for the analysis of the roles of visual contact and leadership instruction in interpersonal synchronization between singers during singing ensemble performances. The protocol relies on the implementation of the novel method of timing detection in ensemble singing presented in the previous chapter, on a two-part piece composed for the study, that was repeated a considerable number of times during the experiment, and on paired sample tests comparing mean and median values across conditions.

The measurement of interpersonal synchronization using the TIMEX algorithm proved to be a valuable method for the study of temporal coordination in ensemble singing, addressing much of the typically laborious manual annotation, and assuring consistency in the analysis. The author’s validation of the extracted timing information from the entire data set was still necessary: some notes that were clearly random mistakes of the singers were not detected as “mistakes” by the algorithm, and therefore had to be removed by hand. In the present study, the number of these wrong notes totalled 1 % of the data set detected automatically by the algorithm. It is likely that the exclusion of such a low percentage of data from the data analysis has not compromised the reliability of the results. For this reason, a manual process that double checked the results and compared the audio performance with the notated score was essential. The implementation of score-audio alignment allowing the detection of the singers’ mistakes is planned for the future.

The singers taking part in the pilot study did not report any issues with the performance of the two-part piece used for the study. However, the presence of several semitones slightly affected the performance of the TIMEX algorithm, requiring manual annotations in the order of the 2 % of the data extracted, which has a slight impact on the objectivity of the automatic detection. This was not evident during the tests evaluating the performance of TIMEX to extract ON, OF, NB and NE, conducted in the previous chapter with a smaller data sample. The Lx clips used for the evaluation of the TIMEX performance were, in fact, around 12 minutes in length, whilst the analysis of synchronization in the pilot study was based on approximately 80 minutes of recordings. It is likely that the small amount of manual processing has not significantly affected the objectivity of the results, and therefore the same score was used for the main study that follows this chapter. This aspect was considered in the second major investigation conducted in this thesis, limiting the use of semitones in the two five-part pieces composed for the study (for more information, see section *6.2.2*).

The music score was sent to the singers well in advance of the experiment, so they could practise the piece; in addition, the singers were allowed to rehearse the score in the 10 minutes before the beginning of the recordings. Nevertheless, both duos preferred to perform from the score on a music stand. The execution of this experiment that involved singers moving often across visual contact condition (i.e., facing each other or turning around) with a music stand was not very practical. This issue was addressed in the design of the main experiment by requesting the mandatory memorisation of the piece prior to the beginning of the recording, as described in the following chapter.

The analysis of the roles of visual contact and leadership instruction was based on an analytical framework consisting of statistical tests comparing mean and median values, in which each value was the result of collapsing part of the dataset, as common in this field of research (Timmers et al. 2014). For example, the *t*-test assessing the impact of visual contact on the precision of NB, compared the 24 asynchrony values representative of the 24 performances sung in the presence of visual contact with the 24 values representative of the performances sung without visual contact. Each value, representative of a repeated performance, was the result of the collapsing across the chosen notes. However, such an analytical framework reduced the analytical power by losing data points. In the main experiment described in the next chapter and in the other two empirical investigations reported in chapter *6* and chapter *7*, these factors were considered. In the remainder of this thesis, multilevel linear models were preferred and implemented, in line with other studies on interpersonal synchronization (Bishop and Goebl 2014; Bishop and Goebl 2015). Multilevel linear modelling increases the statistical reliability, avoiding comparisons between average values and allowing the evaluation of fixed, random and nested effects (Gelman and Hill 2007).

In addition, the analysis of the roles of visual contact and leadership, by pooling together the notes, should be expanded with the analysis of notes that might be particularly crucial to temporal coordination. Synchronization might be greater at the beginning of phonation compared with the synchronization of other notes within a legato phrase. Singers might find it harder to be together when there is no previous temporal reference. Such an approach will provide a more thorough study of synchronization. In the following chapter, this is addressed, and the evaluation is conducted at multiple temporal levels to allow for the analysis of the overall performance as well as crucial notes (see the three stages described in section 5.2.6 for more detail).

The consistency of synchronization in this experiment was quantified by the SD of asynchronies, measured in milliseconds, and by the CV of asynchronies, measured as SD relative to the mean of asynchronies and expressed in %. Although both of these measures are informative of the amount of variation in asynchronies, results expressed in milliseconds (as per the SD measure) might be more relevant to vocal coaches and singing students, who commonly refer to duration in seconds/minutes as absolute time values rather than a ratio (the CV). This was considered in the design of the main study described in the following experiment and also in the second major investigation described in chapter *6*. Thus, the SD measure will be preferred to CV, in the interest of singers and vocal coaches who might benefit from the results of this thesis.

In addition, the potential effect of the considerable number of repeated performances implemented in the design, to allow measurement of the effect of visual contact and leadership, was evaluated. Results showed no visible changes of the synchronisation characteristics across repeated performances that might have been induced by fatigue or learning. In parallel to the evaluation of a protocol, the roles of visual contact and leader-follower relationships were also investigated in the two singing duos. Surprisingly, consistency of NB synchronization as indexed by SD asynchronies was significantly better without visual contact in duo 1. Similarly, precision of NB asynchronies was better in duo 2. Greater consistency of NB synchronization in the absence of visual contact corroborates previous investigations by Keller and Appel (2010) with piano duos analysing the effect of visual contact on CV asynchronies, and might be understood in light of the common requirement to practise with closed eyes and to perform with invisible musicians in recording studios. In addition, the regular rhythmical structure of the piece chosen for the study might have altered, or eliminated, any benefit arising from visual contact between singers. Previous research suggests eye cues are beneficial when irregular musical timings are being performed (Bishop and Goebl 2014). However, the fact that NB synchronization was significantly better without visual contact between singers remains surprising, since previous research has shown no effect of visual contact on median unsigned asynchronies research (Keller and Appel 2010), unlike an apparent benefit from performers not seeing each other. In this condition, another factor, such as audible breathing, might contribute to an extent by boosting interpersonal coordination when visual contact is removed.

Leader-follower relationships also had a significant effect on the consistency of synchronization between singers. When the upper voice led the performance, consistency of NB asynchronies, as indexed by SD and CV values, was significantly worse in duo 1, but better in duo 2, suggesting the need for further investigations. In addition, this study highlighted contrasting effects of the instruction to act as leader or follower on the tendency to precede a co-performer temporally at ON and NB in duo 1, and to precede temporally at NB, NE and OF in duo 2. The designated leader of duo 1 significantly tended to lag the designated follower at ON but precede at NB, when either voice was instructed to lead. But the trend in duo 2 was very different to duo 1. When the upper voice was the leader of duo 2, the leader tended to precede the follower at NB and NE; when the upper voice followed, the leader (i.e. the lower voice) tended to lag at NB and NE, but precede at OF. Furthermore, these contrasting patterns of leader-follower relationships show bidirectional adaptations between performers rather than a clear separation of roles, corroborating previous studies analysing the synchronization of onsets among string quartets (Glowinski et al. 2013; Glowinski et al. 2012; Timmers et al. 2014; Timmers, Endo, and Wing 2013). This also suggests that the reciprocal and iterative adjustments are not limited to the attack of the note, but to the full performance of notes, including NB, NE and OF in addition to ON. In light of this result, it is now of interest to study the relationship between time categories (i.e., ON, NB, NE, and OF); in the next chapter, multilevel linear models will be applied to test the relationships between time categories, as nested effects within the primary fixed effects of visual contact and leadership instruction.

Moreover, the fact that the researcher’s instruction to act as leader or follower of the performance was not associated with the leader showing a significant tendency to precede at note onsets contrasts with findings from piano duos (Goebl and Palmer 2009) whereby musicians assigned the role of leader and asked to play the melody preceded the onsets of those assigned the role of follower performing the accompaniment. These contrasting results might be understood in light of the different materials used in the current experiment, which lacked well-defined melody-accompaniment parts, suggesting that leaders might tend to precede at onsets only when the musical structure facilitates leadership roles.

Finally, the synchronization characteristics observed in these two duos (i.e., average absolute asynchronies of 61 ms for duo 1 and 67 ms for duo 2; average SD asynchronies of 41 ms for both duos; and, median signed asynchronies of -7 ms for duo 1 and -10 ms for duo 2) are in line with the literature analysing synchronization during naturalistic ensemble performance and controlled experiments, reporting typical mean absolute and SD asynchronies falling between 30 and 50 ms, and mean signed asynchronies often close to 0 ms (Keller 2014; Goebl and Palmer 2009; Timmers et al. 2014; Bishop and Goebl 2015; Zamm, Pfordresher, and Palmer 2015). The different music material and/or the different *tempo* used across the literature, and/or the different techniques used to collect the data and extract onsets timing across the studies might have induced these slightly different results.

4.5 Conclusions

The present pilot study has tested a protocol for the study of the impact of the presence/absence of visual contact and leadership instruction on synchronisation between singers during singing duo performances. The protocol relied on the implementation of the new algorithm (TIMEX) described in the previous chapter, on a two-part piece composed for the study that was repeated 48 times during the experiment, and on an analytical framework based on paired *t*-tests and Wilcoxon tests. The measurement of synchronization using TIMEX proved to be effective for 63 % (whilst the remaining 37 % of the data set was based on manual changes), addressing much of the laboriousness typical of the manual annotation. There were no visible changes on the synchronisation characteristics across repeated performances that might have been induced by fatigue or learning. Modifications to the procedure and data analysis to optimize the results have been considered, and will be implemented in the main study described in the following chapter.

The results from the analysis of visual contact and leadership suggests that controlling visual contact and instructing singers to act as leader or follower might significantly affect synchronization between singers. Surprisingly, the consistency and precision of the synchronization of note beginnings increased when there was no visual contact between singers in the case of, respectively, duo 1 and duo 2. Results also revealed patterns of bidirectional adaptations between singers, which were different across duos. These preliminary findings indicate fruitful avenues for the main study that follows.

5 The Roles of Visual Contact and Leadership Instruction in Singing Duo Performances

5.1 Introduction

The previous chapter described a pilot study of the impact of visual contact and leadership instruction on synchronization between singers in two singing duos. The protocol and the implementation of the TIMEX algorithm introduced in chapter *3*, proved to be a successful method for the study of synchronization, and avenues to improve the procedure and the analytical framework were identified. In addition, the pilot study reported precision of synchronization between 41-47 ms across duos, consistency of synchronization of 41 ms for both duos, and the tendency to precede/lag between -4 ms and -10 ms, depending on the duo analysed. Results suggest that these main synchronization characteristics might be affected by the presence/absence of visual contact between singers during singing performances, particularly with regards to note beginnings. Instructing singers to act as leader or follower might also affect the tendency to precede or lag a co-performer and the consistency of synchronization at note beginnings, although in different ways across duos. When the upper voice was assigned the role of leader, consistency of note beginning asynchronies was significantly worse in one duo, but better in the other. Finally, results from the previous preliminary investigation indicated bidirectional temporal adaptation between singers and changes in synchronization according to the time category considered [i.e., onset (ON) and offset (OF) of phonation, and note beginning (NB) and ending (NE) within a legato phrase].

The successful implementation of the protocol and the preliminary findings lead to the main study presented in the current chapter[6], guiding the research questions and hypotheses formulated below. Specifically, the following research questions have been conceived:

---

[6] Most of this chapter, including figures and tables, are reported in D'Amario, Daffern, and Bailes (2018c), and is © the authors, licensed CC-BY.

- What are the main characteristics of synchronization in singing duos, in terms of precision and consistency of synchronization, and tendency to precede/lag a co-performer?
- What are the differences in the synchronization patterns related to different time categories, i.e. ON, OF, NB, and NE?
- Do visual contact and acting as leader or follower affect synchronization between singers in singing duos?
- Are differences in the synchronization patterns by time categories affected by visual contact and leadership instruction and/or associated with the beginning of phonation?

Although this experiment is explorative in nature, it was hypothesised that the presence/absence of visual contact between singers and the instruction to act as leader or follower would affect synchronization during singing duo performances, based on the results of the pilot study. It was also hypothesized that the effects of visual contact and leadership instruction change in relation to note beginnings and endings, as found in the pilot. Finally, it was conjectured that the singing duos would demonstrate a mutual adaptation of roles rather than a clear separation between leader/follower, as found in the previous chapter.

5.2 Method

5.2.1 Participants

Ethical approval for the study (with reference D'Amario151127) was obtained from the Physical Sciences Ethics Committee (PSEC) at the University of York (UK). Twenty-four singing students from the Department of Music at the University of York participated in the current experiment (14 females, age $M = 20.9\ yo$, $SD = 2.9\ yo$). Twenty of them were undergraduate students, and four of them postgraduate students with singing as first study. They had at least three years' formal singing practice ($M = 8.6\ y$, $SD = 4.5\ y$) and at least five years' experience performing in a singing ensemble ($M = 10\ y$, $SD = 5.7\ y$), but they had not sung together prior to the experiment. They reported having normal hearing and not having absolute pitch.

5.2.2 Stimulus

This study made used of the vocal duo exercise that was composed for the pilot study reported in the previous chapter. The piece features a mostly homophonic texture that facilitates the analysis of synchronization, as shown in **Figure 3.5**. For a detailed description of the characteristics of the piece and the notes chosen for the analysis, see section *3.3.1*.

5.2.3 Apparatus

The experiment took place in the same recording studio at the University of York, where the pilot study was conducted (i.e., a studio room treated with absorptive acoustic material). As in the pilot study, audio data were collected using head-mounted close proximity microphones (DPA 4065), placed on the cheek at approximately 2.5 cm from the lips, and a stereo condenser microphone (Rode NT4) placed at equal distance in front of the singers at approximately 1.5 m from the lips. As in the pilot study, electrolaryngograph (Lx) recordings were also collected using Lx electrodes from Laryngograph Ltd. (www.laryngograph.com) placed on the neck. The Lx and acoustic recordings from the head-mounted microphones were used for the analysis of synchronization. Stereo recordings were also collected and used for the perception study reported in chapter *7*. The six outputs (two Lx, two head-mounted microphones, and one stereo microphone comprising right and left channel) were connected to a multichannel hard disk recorder (Tascam DR680) and recorded at a sampling frequency of 48 kHz and 32-bit depth. Compared with the pilot study, a minor difference was introduced regarding the dynamic range of the recorded signal. The bit depth of 24 in the pilot study, which already provided a very high-resolution signal, was increased to 32-bit depth.

5.2.4 Design

A total of four conditions were applied in a randomised order, as follows:

- VC_UpperVoiceL: with visual contact (VC), and upper voice designated leader and lower voice follower (UpperVoiceL)
- VC_UpperVoiceF: with visual contact (VC), and upper voice designated follower and lower voice leader (UpperVoiceF)
- NVC_UpperVoiceL: without visual contact (NVC), and upper voice designated leader and lower voice follower (UpperVoiceL)

- NVC_ UpperVoiceF: without visual contact (NVC), and upper voice designated follower and lower voice leader (UpperVoiceF)[7]

The piece was repeated four times in each condition, and each condition was presented three times. As in the pilot study, this study resulted in a 4 (conditions) × 3 (takes, i.e. repeated performances of each condition) × 4 (repeated performances within each take) design featuring a total of 48 repetitions of the piece per duo.

5.2.5 Procedure

The procedure used for this experiment was very similar to that applied in the pilot study. Singers received written and spoken instructions, and gave written informed consent, at the beginning of the session. Singers first practised the piece together for 10 minutes, singing to the vowel /i/ and listening for 10 seconds to a metronome set at 100 beats per minute (BPM) before starting to rehearse. For consistency with the pilot study, the vowel /i/ was chosen for the performances. During the initial 10-minutes rehearsal, singers were allowed to sing from the score using a musical stand. Due to practical issues in the pilot study with the music stand as noted in the previous chapter, singers performed from memory during the recordings. At the end of the 10 minutes, if the singers were able to perform the piece by memory and without error, the four conditions were then presented; otherwise, they were invited to rehearse for 10 more minutes and then the test was repeated. Once the musicians passed the performance test without error, each singer was assigned the role of leader or follower according to the UpperVoiceL and UpperVoiceF conditions. Thus, in the former condition the upper voice was instructed to act as leader, and the lower voice as follower. These roles were reversed in the UpperVoiceF condition (i.e., the upper voice was instructed to follow, and the lower voice to lead). Signs labelled “Leader” and “Follower” were placed on the floor in front of the participants, to facilitate recalling of their role. Each singer only had one assigned part/musical voice; for example, when a given singer was assigned the upper part, he/she was asked to perform that upper part consistently through the experiment. Musicians faced each other at a distance of 1.5 m in the visual condition and turned away from each other at the

---

same distance in the non-visual contact condition. Participants were asked to sing at performance level and were unaware of the purpose of the study.

5.2.6 Analysis

Asynchronies were detected as in the pilot study. Two sets of data including the audio waveform from the head-mounted microphones and the Lx waveform were first imported into Praat (Boersma 2001; Boersma and Weenink 2013) as .wav files, and then fundamental frequency (f_o) estimates were extracted with a time step of 1 ms. The analysis of interpersonal synchronization was conducted on relevant notes, as shown in **Figure 3.5**. For each chosen note, a true starting and ending time stamp value was detected, based on the definition of the four-time categories (i.e., ON, NB, NE, and OF) provided in chapter 3. The extraction of the time categories was automated through the application of the TIMEX algorithm, described and tested in chapter 3 and chapter 4. As in the pilot study, this event detection method was aurally and visually cross-validated for the entire data set by the first author. The Lx signal was scrutinized for soft phonation, which might have not been picked-up by the Lx if the vocal fold oscillation was too soft (Sten Ternström and Johan Sundberg, personal communication, June 2017). But, as in the pilot study, cases of soft phonation were very scarce. Pitch errors due to the musicians singing wrong notes (i.e., entering or delaying the notes for more than 50 % of their values) were analysed comparing the f_o values extracted and the acoustics and Lx recordings with the notated score. Notes in which a timing error occurred were excluded from the analysis. The timing error rate was less than 1 %. As in the pilot study, a manual correction was sometimes required to detect accurate timing information in relation to semitones, when the algorithm failed to detect. The manual annotation regarding semitones that the algorithm failed to detect was approximately 5 %. Furthermore, manual changes for an additional 34 % of the data set were also necessary in cases in which the singers performed a vibrato (within the same note) with an extent larger than a semitone; in these cases, the algorithm identified the frequency oscillations of the vibrato as note beginnings/endings, failing to detect the vibrato. The overall manual changes related to errors associated with timing, semitones, and vibratos were around 40 %.

Interpersonal asynchronies were then calculated as in the pilot study, subtracting the follower's timestamp values from the leader's (leader minus follower) regarding each time category of the selected notes. Asynchronies that fell outside three times the interquartile range (IQR) were automatically identified as extreme outliers through SPSS (IBM SPSS Statistics v. 24) and excluded. The identification of outliers was run for each time category, performance condition and duo.

Three measures of synchronization were investigated, namely:

- Precision of temporal synchronization, as quantified by the absolute asynchronies, and expressed in milliseconds.
- Consistency, as measured by the standard deviation (SD) of absolute asynchronies, and expressed in milliseconds. This has been computed for each time category, note, condition, and duo, across the repeated performances within each condition. For example, SD asynchrony was computed for the onset of note 1 in Duo 1 regarding the VC_UpperVoiceL condition across the 12 repeated performances featuring this time category/note/duo/condition.
- Tendency to precede or lag a co-performer, as quantified by the signed asynchronies, and expressed in milliseconds.

As explained in the previous chapter, the coefficient of variation was not computed in the interests of a key target audience for this thesis (singers and vocal coaches). To increase the practical relevance of the results for the target population, the consistency of synchronization is expressed in milliseconds, rather than as a standard deviation relative to the sample mean expressed in %.

In light of the recommendations highlighted in the pilot study, to understand whether visual contact or leadership had an effect on synchronization, and whether the effects, if any, depend on important voice entry points and/or time category, the analytical framework was deeper than that implemented in the pilot study. Here the analysis was run across the following three stages and levels:

- Stage 1 - High level: considers the effect of the independent variables on the synchronization measures, incorporating the full data set.
- Stage 2 - Medium level: investigates the effect of the independent variables at singers' simultaneous entries, based on the subset of data including note 1, 3, 19, and 22. Notes 1 and 19 were chosen as being points of simultaneous voice entry; whilst note

3 was selected to investigate whether any effect regarding the simultaneous entry at the beginning of the piece disappeared by the next downbeat (i.e., the third downbeat of bar 1, since note 1 is a dotted quarter); for similarity with bar 1, note 22 has been selected, being the third downbeat of that bar as well.

- Stage 3 - Low level: analyses the effect of visual contact and leadership on the time category of those notes where a main effect was found at the medium level. The analysis at this level was conducted to understand whether the effect observed at the medium level, if any, would relate to the beginning of phonation.

In light of the recommendations resulting from the pilot study, stepwise multilevel linear models were chosen to strength the statistical reliability of the fixed effects analyses by providing an evaluation of inter-participant, inter-time category, and inter-note variation (Gelman and Hill 2007). Multilevel linear models were developed for each stage of the analysis (i.e., stage 1-3), response variable (i.e., absolute asynchronies, signed asynchronies, and SD of absolute asynchronies,), and primary fixed factors (i.e., visual contact[8] and leadership), as shown in **Table 5.1**, **Table 5.2**, **Table 5.3**, **Table 5.4**, and **Table 5.5**. Time category and note were also entered in the model as fixed effects nested in the primary fixed effects, or as random effects, depending on the level of the analysis. Participants were treated as a random variable across levels. At the high level of the analysis, models were designed to test the fixed effects of visual contact, leadership, and time category (the latter nested within the two former), and the random effects of participant and note. At the medium level, models tested the fixed effects of visual contact, leadership, and note subset, i.e. note 1, 3, 19, and 22 (note subset nested within the two former) and the random effect of participants. At the low level, models were developed to investigate the fixed effects of visual contact, leadership, and time category (the latter nested within the two former), and the random effect of participants. The models were implemented in R Studio (RStudio 2015) using the lme4 package.

---

[8] In other words, leadership instruction and visual contact conditions were inputted as independent variables/primary fixed factors. In light of the process followed for the computation of the signed asynchronies (designated leader minus designated follower), the measured asynchronies computed by leadership instruction were treated as dependent variables/response variables.

The investigation was first conducted at the high level, then the analysis of each response variable proceeded at a medium level when a significant fixed effect was found. Similarly, the analysis moved to the low level if a significant fixed effect was found at the medium level. Conversely, if a significant effect was not found at a high or medium level, the analysis was not carried over to a deeper level (i.e., from high to medium, or from medium to low).

A Bonferroni correction was implemented to reduce the possibility of obtaining spurious significant results from multiple multilevel linear models. The alpha level was set at $p = 0.0028$, based on the assumption that a total of 18 models might have been developed [3 (stages) × 3 (response variables) × 2 (primary fixed factors)], if the analyses proceeded from stage 1 to stage 3.

5.3 Results

An initial overview of the full data set is provided in section *5.3.1* by way of descriptive statistics, with the purpose to scrutinize the main characteristics of synchronization in singing ensembles, irrespective of condition (i.e., visual contact, and the instruction to act as leader or follower). The remaining two sections (see *5.3.2* and *5.3.3*) present the results of the analyses of the main effects of visual contact and the instruction to act as leader or follower on interpersonal synchronization, respectively. β - fixed effect coefficients on the predictor being considered - are given below and in **Table 5.1**, **Table 5.2**, **Table 5.3**, **Table 5.4**, and **Table 5.5** with reference to the specified base level of the factor, i.e. NB, NE or OF versus the base level ON, and note 3, 19 or 22 versus the base level note 1. The β coefficients indicate that for each 1 unit increase in the predictor being considered, the effect of the given predictor changes by the amount specified by the β coefficient.

5.3.1 Synchronization Characteristics

The analysis of the overall synchronization was computed irrespective of performance condition and time categories, taking all notes together and averaging for each duo. Results show that the precision of overall synchronization computed on the mean of absolute asynchronies was on average 58.99 ms $(SD = 11.13\,ms)$, consistency indexed by SD of absolute asynchronies was 67.06 ms $(SD = 11.85\,ms)$, whilst the tendency to precede/lag as indexed by the median of signed asynchronies was -4.06 ms $(IQR = 4.38\,ms)$.

The full sample data were scrutinized to investigate changes in the asynchronies across the course of the 48 repeated performances, by averaging the asynchronies for each measure (mean, SD, and median) and each performance across the 12 duos. **Figure 5.1** represents these data and suggests that, with practice, there was no discernible improvement in synchronization between the singers.

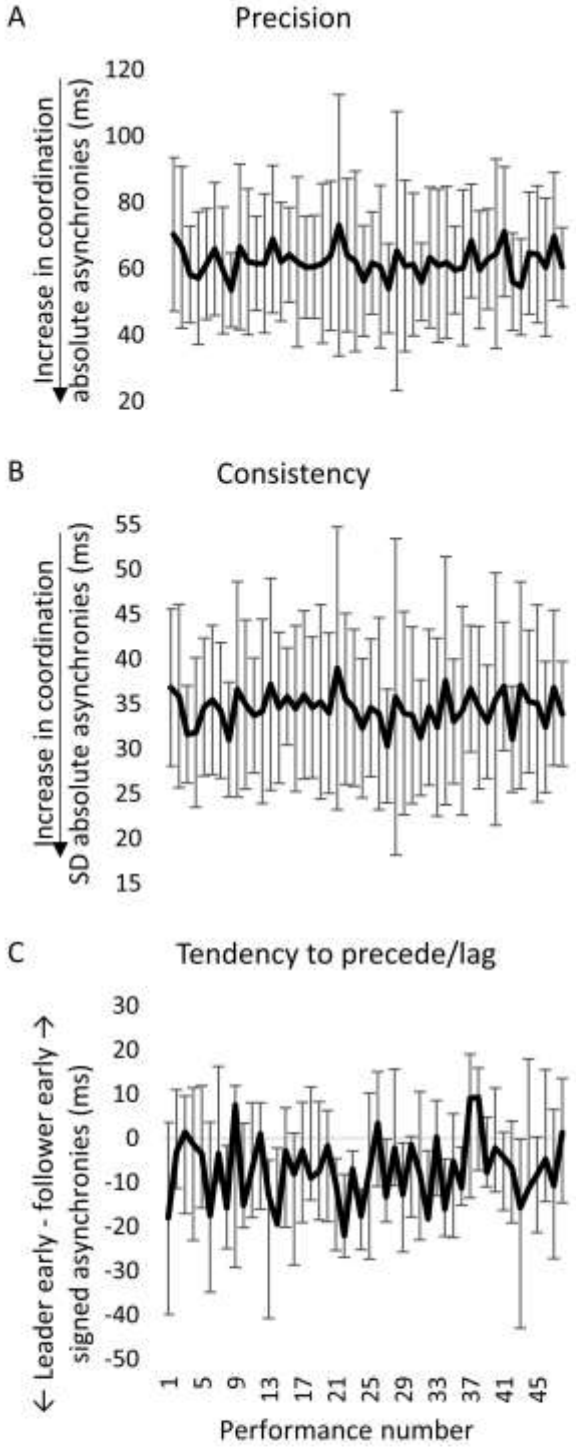


Figure 5.1 *Performance across the 48 trials, indicating precision of synchronization in panel A (mean of absolute asynchronies), consistency in panel B (SD of absolute asynchronies), and tendency to precede or lag a co-performer in panel C (median of signed asynchronies). Error bars of precision and consistency represent standard error of the mean, whilst error bars of the tendency to precede indicate interquartile range of the median.*

5.3.2 Visual Contact Effect

5.3.2.1 Precision

The analysis conducted at stage 1, based on the multilevel linear model developed as explained above, demonstrated that the presence/absence of visual contact between singers predicted precision in the synchronization $[\beta(-31.7), t\ (25000) = -10.5, p < 0.001]$ (see **Table 5.1**, row 1). As shown in **Figure 5.2A**, precision of synchronization was significantly better when visual contact between singers was present $(M = 56.0\ ms, SD = 48.2\ ms)$, compared with when visual contact was absent $(M = 60.1\ ms, SD = 53.6\ ms)$. The variance partition coefficient (VPC) among participants and notes was 0.027 and 0.043, which indicates that only 2.7 % and 4.3 % of the variability of the effect of visual contact can be attributed to participants and chosen note, respectively. In the presence of visual contact between singers, precision temporally computed at ON was better than that computed at NE $[\beta(8.8), t\ (22587) = 3.3, p < 0.001]$, and OF $[\beta(19.4), t\ (20330) = 5.7, p < 0.001]$ (see **Table 5.1**, rows 2-4). Interestingly, when visual contact between singers was absent, the relationship between time categories changed: precision computed at ON was lower than that computed at NB $[\beta(-21.9), t\ (20893) = -8.0, p < 0.001]$, and NE $[\beta(-21.8), t\ (22537) = -8.3, p < 0.001]$ (see **Figure 5.2B**, and **Table 5.1**, rows 5-7). *Post-hoc* tests between same pairs of time categories (e.g., ON in presence and absence of visual contact), calculated with Holm correction for multiple comparisons, show that precision of NB synchronization was significantly better in the presence of visual contact $(M = 54.0\ ms, SD = 48.1\ ms)$, than in its absence $[M = 58.0\ ms, SD = 50.2\ ms, t = 4.7, p < 0.001]$; likewise, precision in the synchronization computed at ON was better with visual contact between singers $(M = 51.7\ ms, SD = 49.0\ ms)$, than without $[M = 83.2\ ms, SD = 92.0\ ms, t = 10.5, p < 0.001]$.

When the effect of visual contact was investigated in relation to notes 1, 3, 19, and 22 (i.e., medium level of the analysis), results show that, in the presence of visual contact, precision at note 1 was significantly greater than that computed at note 19 $[\beta(-20.8), t\ (4247) = -6.2, p < 0.001]$, and note 22 $[\beta(-19.3), t\ (4247) = -5.8, p < 0.001]$ (see **Table 5.1**, rows 9-11). When visual contact was absent, the coefficients of these relationships were even larger: synchronization at note 1 was greater than that at note 3 $[\beta(-27.2), t\ (4247) = -8.2, p < 0.001]$, note 19 $[\beta(-37.6), t\ (4247) = -11.3, p <$

$0.001]$, and note 22 $[\beta(-40.2), t\ (4247) = -12.1, p < 0.001]$ (see **Table 5.1**, rows 12-14). The variability of this effect among participants was small $(VPC = 4.7\ \%)$. *Post-hoc* comparisons demonstrate that this effect was associated with note 1: precision of synchronization was significantly better with visual contact $(M = 66.9\ ms, SD = 55.8\ ms)$, compared to without $[M = 90.9\ ms, SD = 91.3\ ms, t = 7.139, p < 0.001]$, as shown in **Figure 5.2C**.

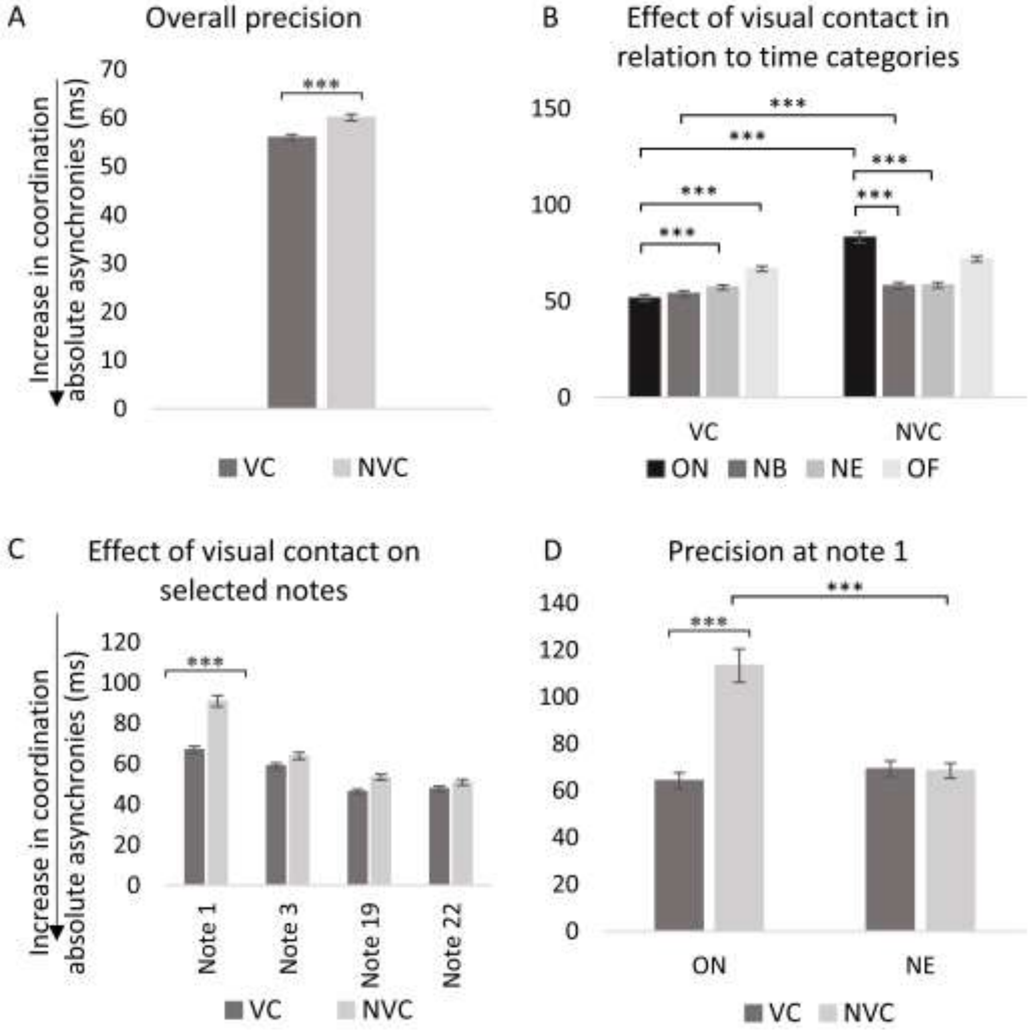


Figure 5.2 *Panel A: effect of visual contact on precision of synchronization computed overall; panel B: effect of visual contact in relation to time categories; panel C: effect of visual contact computed on a subset of notes, but only comparison between the same pairs of notes is presented; panel D: effect of visual contact on note 1. Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $*** \ p < 0.001$.

The analysis conducted at stage 3 highlighted that without visual contact, precision at ON was significantly greater than that at NE $[\beta(-44.6), t\ (1035.8) = 7.4, p < 0.001]$ (see **Table**

5.1, row 17), and that precision at ON was better with visual contact between singers $(M = 64.2\,ms, SD = 56.3\,ms)$, than without $[(M = 113.4\,ms, SD = 114.7\,ms), t = 8.0, p < 0.001]$ (see **Figure 5.2D**). The variability of this effect among subjects was small ($VPC = 4.7$ %).

In summary, these findings show that the presence/absence of visual contact predicted the precision of synchronization, which was better when the visual contact between singers was present, compared with when the visual contact was absent. This effect was constant among participants and was associated with the onset of phonation at the beginning of the piece.

Table 5.1 *Overview of the multilevel linear models developed to investigate the precision of synchronization, with primary effects of visual contact, nested effects of time category and crucial notes, and the random effects of participants and chosen notes.*

Stage of analysis	**Fixed effects**	**β coefficients and significance**			**Random effects**	**Row number**
Stage1: Overall	Visual contact	$\beta(-31.7) ***, t\ (25000) = -10.5$			Participants	1
	Time category nested	VC	NB	n.s.	Chosen notes	2
			NE	$\beta(8.8) ***, t\ (22587) = 3.3$		3
			OF	$\beta(19.4) ***, t\ (20330) = 5.7$		4
		NVC	NB	$\beta(-21.9) ***, t\ (20893) = -8$		5
			NE	$\beta(-21.8) ***, t\ (22537) = -8.3$		6
			OF	n.s.		7
Stage 2: Notes subset	Visual contact	$\beta(-23.9) ***, t\ (24250) = -7.1$			Participants	8
	Crucial notes nested	VC	3	n.s.		9
			19	$\beta(-20.8) ***, t\ (4247) = -6.2$		10
			22	$\beta(-19.3) ***, t\ (4247) = -5.8$		11
		NVC	3	$\beta(-27.2) ***, t\ (4247) = -8.2$		12
			19	$\beta(-37.6) ***, t\ (4247) = -11.3$		13
			22	$\beta(-40.2) ***, t\ (4247) = -12.1$		14
Stage 3: Significant note	Visual contact	n.s.			Participants	15
	Time category nested	VC	NE	n.s.		16
		NVC	NE	$\beta(-44.6) ***, t\ (1035.8) = 7.4$		17

n.s. = not significant; $*** p < 0.001$. β - fixed effect coefficients on the predictor being considered - are given above with reference to the specified base level of the factor, i.e. NB, NE or OF versus the base level ON, and note 3, 19 or 22 versus the base level note 1.

5.3.2.2 Consistency

The analysis conducted at the high level demonstrates that the presence/absence of visual contact predicted the consistency of synchronization as indexed by the SD of absolute asynchronies $[\beta(-19.6), t(2224) = -6.1, p < 0.001]$ (see **Table 5.2**, row 1). Synchronization was more consistent with visual contact between singers $(M = 38.2\, ms, SD = 17.1\, ms)$ than without $(M = 41.9\, ms, SD = 18.7\, ms)$, as shown in **Figure 5.3A**. The variability of this effect among participants and chosen notes was small, $VPC = 9.3\ \%$ and $VPC = 14\ \%$. With visual contact, synchronization temporally computed at ON was more consistent than that at OF $[\beta(12.2), t(2065.3) = 3.5, p < 0.001]$ (see **Table 5.2**, row 4). However, when visual contact was absent, the relationships between time categories changed again: synchronization computed at ON was less consistent than that at NB $[\beta(-13.8), t(2106.8) = -4.8, p < 0.001]$, and NE $[\beta(-12.5), t(2181.6) = -4.5, p < 0.001]$ (see **Table 5.2**, rows 5-7). As highlighted by *post-hoc* testing, Holm corrected for multiple comparisons, synchronization temporally calculated at ON was more consistent in the presence of visual contact $(M = 35.8\, ms, SD = 17.2\, ms)$, than without $[M = 55.2\, ms, SD = 41.7\, ms, t = 6.1, p < 0.001]$, as shown in **Figure 5.3B**.

Further analysis conducted on notes 1, 3, 19 and 22 (medium level of analysis), demonstrates that in the presence of visual contact, synchronization temporally computed at note 1 was less consistent than that at note 19 (the second simultaneous voice entry of the piece) $[\beta(-12.7), t(370.1) = -3.2, p < 0.01]$, as shown in **Figure 5.3C** and **Table 5.2**, row 8. The relationships between this subset of notes were affected by the absence of visual contact between singers: synchronization at note 1 was even less consistent than that at note 19 $[\beta(-20.9), t(370) = -5.4, p < 0.001]$, and note 22 $[\beta(-19.9), t(370.1) = -3.2, p < 0.001]$ (see **Table 5.2**, rows 13-14). The variability of this effect among participants was $VPC = 13\ \%$. *Post-hoc* comparisons between same pairs of notes in the two different conditions show that this effect relied on the first note of the piece, $t = -3.4, p < 0.05$. Synchronization between singers computed at note 1 was more consistent with visual contact between singers $(M = 45.0\, ms, SD = 16.7\, ms)$, than without $(M = 58.1\, ms, SD = 40.0\, ms)$.

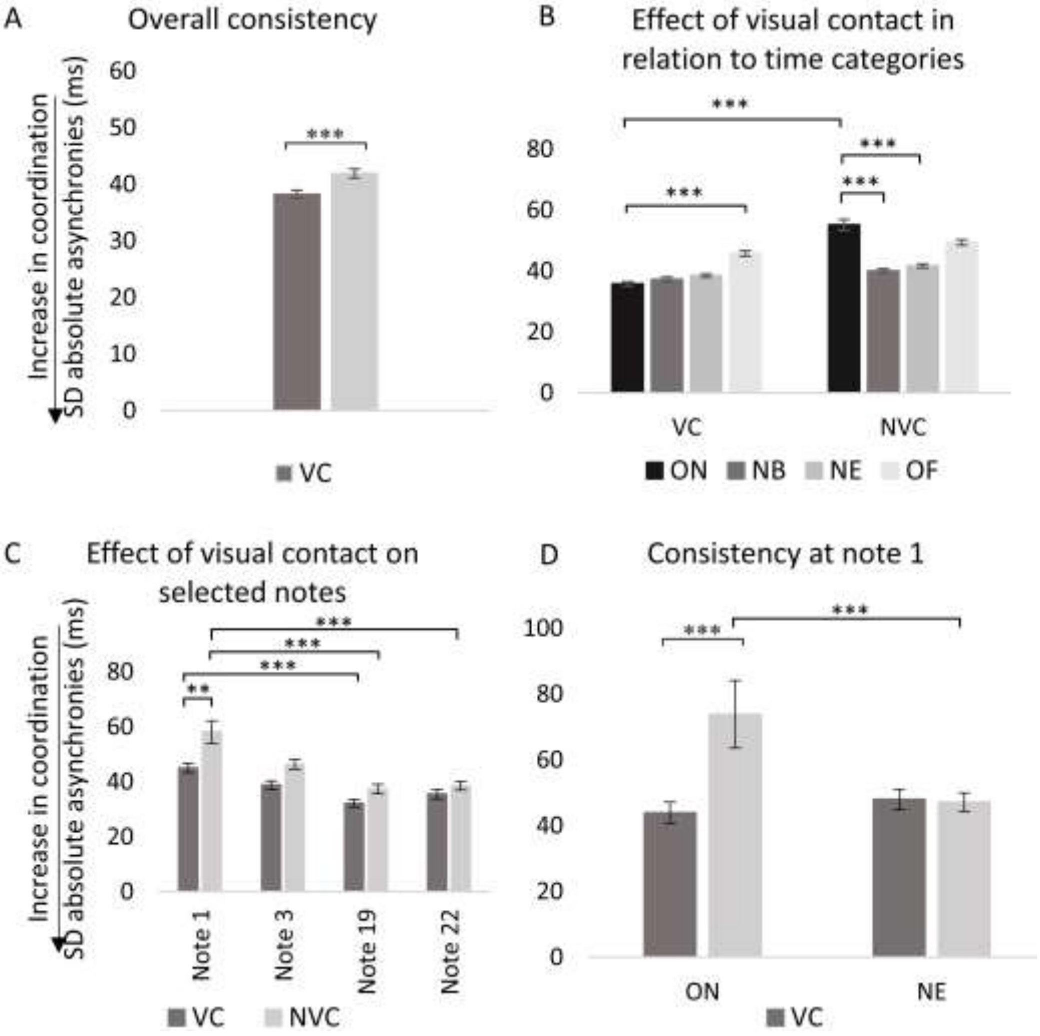


Figure 5.3 *Effect of visual contact on consistency of synchronization, as indexed by SD of absolute asynchronies, computed: for the overall piece (panel A), in relation to time categories (panel B), to a subset of notes (panel C), and to note 1 (panel D). Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $** \, p < 0.01$,$*** \, p < 0.001$.

The analysis which focused on the first note of the piece demonstrated that without visual contact, consistency at ON was significantly greater than that at NE $[\beta(-26.7), t\ (81) = 3.8, p < 0.001]$ (see **Table 5.2**, rows 17), and that consistency at ON was better with visual contact between singers $(M = 43.9\ ms, SD = 15.6\ ms)$, than without $[M = 73.8\ ms, SD = 50.4\ ms, t = 4.2, p < 0.001]$ (see **Figure 5.3D**). The variability of this effect among participants was small, $VPC = 26.7\ \%$.

In summary, the presence and absence of visual contact had a significant effect on the consistency of the temporal coordination of the overall piece: synchronization was more consistent with visual contact between singers, than without it. This effect was consistent among participants and was associated with the synchronization of the onset of the first note of the piece.

Table 5.2 *Overview of the multilevel linear models developed to investigate the consistency of synchronization, with primary effects of visual contact, nested effects of time category and crucial notes, and the random effects of chosen notes and participants.*

Stage of analysis	Fixed effects	β coefficients and significance			Random effects	Row number
Stage1: Overall	Visual contact	$\beta(-19.6)\ ***, t(2224) = -6$			Participants	1
	Time category nested	VC	NB	n.s.	Chosen notes	2
			NE	n.s.		3
			OF	$\beta(12.3)\ ***, t(2065) = 3.5$		4
		NVC	NB	$\beta(-13.8)\ ***, t(2107) = -4.8$		5
			NE	$\beta(-12.5)\ ***, t(2182) = -4.5$		6
			OF	n.s.		7
Stage 2: Notes subset	Visual contact	$\beta(-13.5)\ ***, t(370) = -3.4$			Participants	8
	Crucial notes nested	VC	3	n.s.		9
			19	$\beta(-12.7)\ **, t(370) = -3.2$		10
			22	n.s.		11
		NVC	3	n.s.		12
			19	$\beta(-20.9)\ ***, t(370) = -5.4$		13
			22	$\beta(-19.9)\ ***, t(370.1) = -3.2$		14
Stage 3: Significant note	Visual contact	n.s.			Participants	15
	Time category nested	VC	NE	n.s.		16
		NVC	NE	$\beta(-26.7)\ ***, t\ (81) = 3.8$		17

n.s. = not significant; $**\ p < 0.01$; $***\ p < 0.001$. β - fixed effect coefficients on the predictor being considered - are given above with reference to the specified base level of the factor, i.e. NB, NE or OF versus the base level ON, and note 3, 19 or 22 versus the base level note 1.

5.3.2.3 Tendency to Precede or Lag a Co-Performer

The presence/absence of visual contact between singers predicted the tendency to precede or lag a co-performer $[\beta(28.3), t(25091) = 5.9, p < 0.001]$ (see **Table 5.3**, row 1). The variability of this result attributed to the participants is 0.09 %, and the variability among the chosen notes is 0.02 %. One sample t-tests conducted for difference from 0 show that the designated leader significantly tended to be ahead of the co-performer in the presence of visual contact $[t(12491) = -3.7, p < 0.001]$, as well as without $[t(12661) = -12.0, p < 0.001]$, as shown in **Figure 5.4A**. However, the amount by which the leader tended to precede the co-performer without visual contact $(M = -8.7\,ms, SD = 81.8\,ms)$ was greater than with $(M = -2.5\,ms, SD = 75.1\,ms)$, as highlighted by the fixed-effect coefficient above. In addition, with visual contact, the amount of leading observed at ON was greater than that at NB $[\beta(16.2), t(1264) = 4.0, p < 0.001]$, NE $[\beta(18.0), t(2031) = 4.6, p < 0.001]$, and OF $[\beta(23.8), t(1310) = 4.8, p < 0.001]$. Without visual contact, those relationships are amplified, as highlighted by the following fixed effects coefficients: the amount of leading found at ON was even greater than that observed at NB $[\beta(41.3), t(1247) = 10.3, p < 0.001]$, NE $[\beta(40.1), t(1997) = 10.2, p < 0.001]$, and OF $[\beta(52.4), t(1295) = 10.7, p < 0.001]$. *Post-hoc* testing between the same pairs of time categories, correcting using the Holm method for multiple comparisons, demonstrates that these effects were associated with the tendency to precede/lag a co-performer at ON. The amount of leading computed at ON when visual contact was absent $(M = -48.2\,ms, SD = 115.4\,ms)$, was significantly greater than that observed when visual contact was present $[M = -18.9\,ms, SD = 70.3\,ms, t = -5.9, p < 0.001]$ (see **Figure 5.4B**).

The analysis of the tendency to precede/lag a co-performer in the presence of visual contact demonstrated that the subset of notes was not a predictor of the tendency to precede/lag (see **Table 5.3**, rows 9-11). Conversely, when visual contact was absent, the amount of leading observed at note 1, was significantly greater than that computed at note 3 $[\beta(29.8), t(4261) = 6.0, p < 0.001]$, note 19 $[\beta(37.8), t(4261) = 7.6, p < 0.001]$, and note 22 $[\beta(44.1), t(4261) = 8.9, p < 0.001]$, as shown in **Figure 5.4C** and **Table 5.3**, rows 12-14. The variability of this effect among participants was small $(VPC = 1.3\ \%)$. *Post-hoc* comparisons between same pairs of notes in the two different conditions demonstrate that the effect of presence/absence of visual contact between singers is associated with the synchronization of note 1; the amount of leading was significantly greater when visual contact was absent $(M = -47.0\,ms, SD = 121.3\,ms)$, compared with when visual contact was present $[M = -14.7\,ms, SD = 88.9\,ms, t = -6.4,\ p < .001]$.

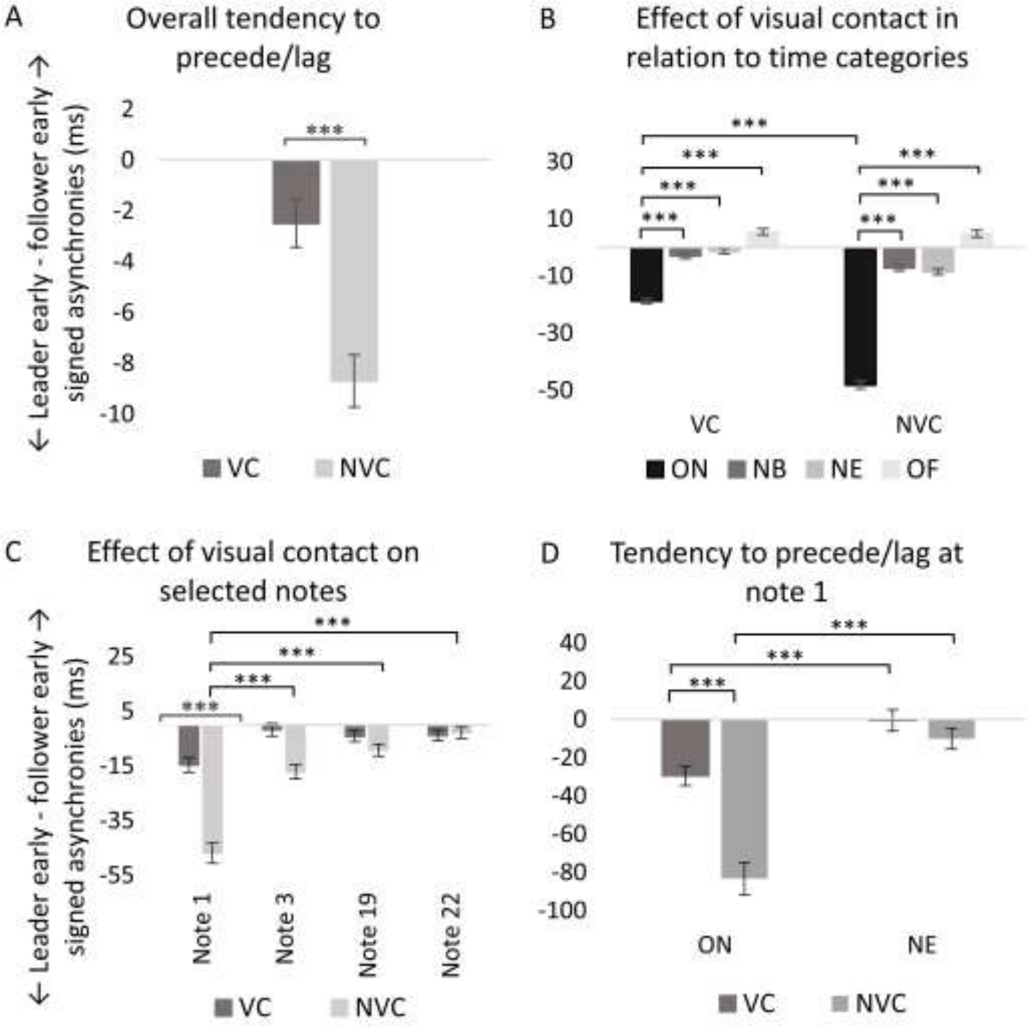


Figure 5.4 *Panel A: effect of visual contact on tendency to precede/lag a co-performer computed overall; panel B: effect of visual contact in relation to time categories; panel C: effect of visual contact computed on a subset of notes; panel D: effect of visual contact on note 1. Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $*** \; p < 0.001$.

The analysis of the first note of the piece demonstrated that the time category predicted the tendency to precede/lag a co-performer in both conditions, i.e. with and without visual contact (see **Table 5.3**, rows 16-17). However, the amount of leading was greater in the absence of visual contact, as highlighted by the fixed-effect coefficients: the tendency to precede/lag the co-performer at ON was larger than that at NE in the presence of visual contact $[\beta(29.1), t(1049.8) = 3.4, p < 0.001]$; but, the leading observed at ON was even greater than that at NE without visual contact $[\beta(73), t(1049.9) = 8.5, p < 0.001]$. ***Post-hoc*** testing highlighted that these effects were associated with the ON: the tendency to precede/lag the co-performer was greater when visual contact was absent $(M = -83.3\,ms, SD = 138.3\,ms)$, compared with when visual contact was kept $[M = -29.6\,ms, SD = 82.7\,ms, t = -6.2, p < 0.001]$ (see **Figure 5.4D**). The variability among participants of the effect of visual contact on the tendency to precede/lag note 1 was small $(VRP = 7.2\,\%)$.

In summary, these results demonstrate that without visual contact, the designated leader showed overall a stronger tendency to precede the designated follower, than with visual contact. This effect was consistent among participants and was associated with a stronger tendency to precede the designated follower at the onset of note 1 when no visual contact was available.

Table 5.3 *Overview of the multilevel linear models developed to investigate the tendency to lead/lag synchronization, with the primary effects of visual contact, nested effects of time category and crucial notes, and the random effects of participants and chosen notes.*

Stage of analysis	Fixed effects	β coefficients and significance			Random effects	Row number
Stage1: Overall	Visual contact	$\beta(28.3) ***, t(25091) = 5.9$			Participants	1
	Time category nested	VC	NB	$\beta(16.2) ***, t(1264) = 4$	Chosen notes	2
			NE	$\beta(18) ***, t(2031) = 4.6$		3
			OF	$\beta(23.8) ***, t(1310) = 4.8$		4
		NVC	NB	$\beta(41.3) ***, t(1247) = 10.3$		5
			NE	$\beta(40.1) ***, t(1997) = 10.2$		6
			OF	$\beta(52.4) ***, t(1295) = 10.7$		7
Stage 2: Notes subset	Visual contact	$\beta(32.1) ***, t(4270) = -1$			Participants	8
	Crucial notes nested	VC	3	n.s.		9
			19	n.s.		10
			22	n.s.		11
		NVC	3	$\beta(29.8) ***, t(4261) = 6$		12
			19	$\beta(37.8) ***, t(4261) = 7.6$		13
			22	$\beta(44.1) ***, t(4261) = 8.9$		14
Stage 3: Significant note	Visual contact	n.s.			Participants	15
	Time category nested	VC	NE	$\beta(29.1) ***, t(1049.8) = 3.4$		16
		NVC	NE	$\beta(73) ***, t(1049.9) = 8.5$		17

n.s. = not significant; $*** p < 0.001$. β - fixed effect coefficients on the predictor being considered - are given above with reference to the specified base level of the factor, i.e. NB, NE or OF versus the base level ON, and note 3, 19 or 22 versus the base level note 1.

5.3.3 Effect of the Instruction to Act as Leader or Follower

5.3.3.1 Precision and Consistency

The instruction to act as leader or follower of the performance did not predict precision of the synchronization of the whole piece, as shown in **Figure 5.5A** and **Table 5.4**, row 1. This result did not vary greatly among participants ($VPC = 2.6\ \%$) or note ($VPC = 4.3\ \%$). Precision at ON was significantly greater compared with NB when the upper voice was instructed to lead $[\beta(-8.9), t(20806) = -3.2, p < 0.01]$, and also when instructed to follow $[\beta(-7.4), t(20926) = -2.7, p < 0.01]$ (see **Figure 5.5B** and **Table 5.4**, rows 2 and 5). When the upper voice was leading, for each unit of increase in the precision computed at onsets, precision at note beginnings decreased by 8.9 ms; when the upper voice was lagging, precision at note beginnings decreased by 7.4 ms. *Post-hoc* tests did not show a significant difference between same pairs of time categories in the two different conditions (i.e., when upper voice was instructed to lead or follow). Since the leadership instruction was not a significant predictor of precision at stage 1, the analysis was not conducted for deeper levels, i.e. stage 2 and 3.

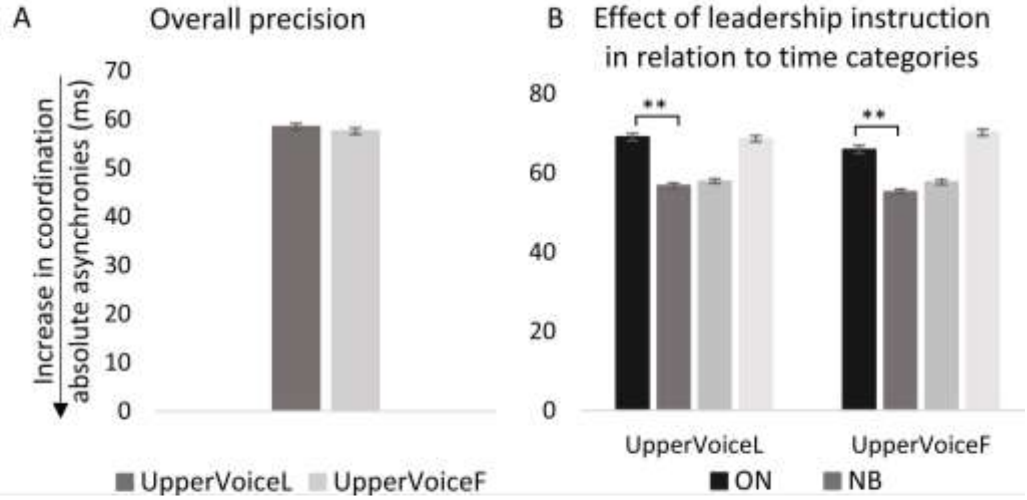


Figure 5.5 *Effect of leadership instruction on the precision of synchronization computed for the whole piece (panel A), and in relation to the time categories (panel B). Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $** \ p < 0.01$.

As shown in **Figure 5.6A** and **Table 5.4**, row 8, the effects of instruction on the consistency of synchronization as indexed by the SD of absolute asynchronies were not found, and the variability of these results was small among participants (for SD asynchronies, $VPC = 9.1\ \%$)

and notes (for SD asynchronies, $VPC = 13.6\ \%$). The instruction was not associated with differences between ON and NB, ON and NE, or ON and OF (see **Figure 5.6B** and **Table 5.4**, row 9).

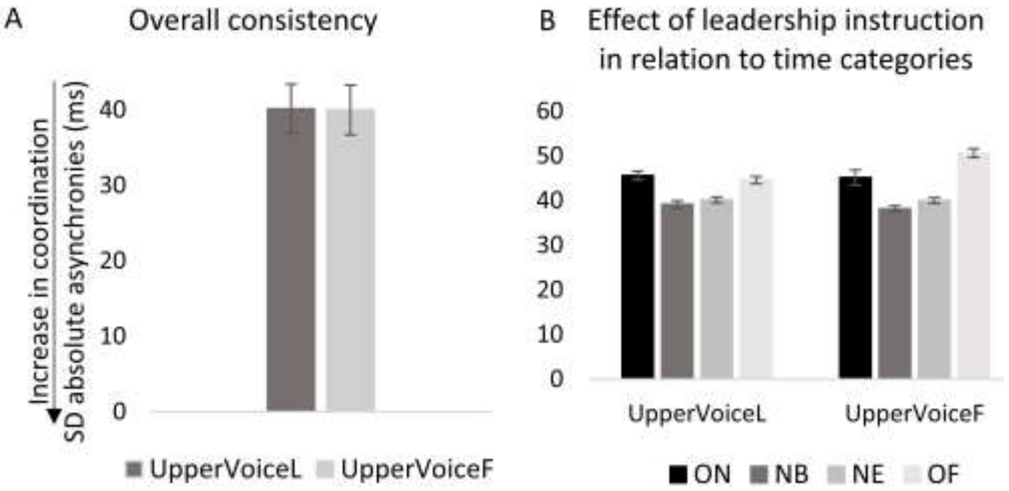


Figure 5.6 *Effect of leadership instruction on the consistency of synchronization as indexed by SD of absolute asynchronies, computed for the overall piece (panel A), and in relation to time categories (panel B).*

Table 5.4 *Overview of the multilevel linear models developed to investigate the precision and consistency of synchronization, with the primary effects of leadership instruction, nested effects of time category, and the random effects of participants and chosen notes.*

Synchronization parameter	Stage of analysis	Fixed effects	β coefficients and significance			Random effects	Row number
Precision	Stage1: Overall	Leadership	n.s.			Participants	1
		Time category	UVL	NB	$\beta(-8.9) **, t(20806) = -3.2$	Chosen notes	2
				NE	n.s.		3
				OF	n.s.		4
			UVF	NB	$\beta(-7.4) **, t(20926) = -2.7$		5
				NE	n.s.		6
				OF	n.s.		7
Consistency	Stage1: Overall	Leadership	n.s.			Participants	8
		Time category	n.s.			Chosen notes	9

n.s. $=$ not significant; $** p < 0.01$. β - fixed effect coefficients on the predictor being considered - are given above with reference to the specified base level of the factor, i.e. NB, NE or OF versus the base level ON.

5.3.3.2 Tendency to Precede or Lag a Co-Performer

The analysis conducted at stage 1 shows that the instruction to act as leader or follower of the performance predicted the tendency to precede/lag a co-performer $[\beta(-14.3), t(25112) = -3.0, p < 0.01]$ (see **Figure 5.7A** and **Table 5.5**, row 1); the variability of this effect among participants was 0.1 % and chosen notes was 0.2 %. One sample t-tests conducted for difference from 0 show that the designated leader significantly tended to be ahead of the co-performer when the upper voice was instructed to lead $[M = -10.5\,ms, SD = 78.1\,ms, t(12491) = -3.7, p < 0.001]$. When the upper voice was instructed to follow, nobody tended to precede/lag the co-performer overall. In addition, the tendency to precede/lag changes according to the time category regardless of the instruction (see **Figure 5.7B**). When the upper voice was instructed to lead, the degree of leading observed at ON was greater than that found at NB $[\beta(26.2), t(1247) = 6.6, p < 0.001]$, NE $[\beta(37.3), t(1996) = 9.5, p < 0.001]$, and OF $[\beta(43.9), t(1286) = 9.1, p < 0.001]$ (see **Table 5.5**, rows 2-4). Similarly, when the upper voice was instructed to follow, the amount of leading by the lower voice observed at ON was greater than that found at NB $[\beta(31.5), t(1271) = 7.9, p < 0.001]$, NE $[\beta(20.8), t(2044) = 5.3, p < 0.001]$, and OF $[\beta(32.4), t(1327) = 6.6, p < 0.001]$, see **Table 5.4**, rows 5-7.

The analysis of the tendency to precede/lag a co-performer demonstrated that, when the upper voice was instructed to lead, the amount of leading observed at note 1 was significantly greater than that computed at note 3 $[\beta(15.1), t(4260) = 3.0, p < 0.001]$, and note 22 $[\beta(20.7), t(4260) = 4.2, p < 0.001]$, as shown in **Figure 5.4C** and **Table 5.5**, rows 9-11. When the upper voice was instructed to follow, those relationships are amplified, as highlighted by the following fixed effects coefficients: the amount of leading found at note 1 was even greater than that observed at note 3 $[\beta(27.7), t(4260) = 5.5, p < 0.001]$ and note 22 $[\beta(34.4), t(4260) = 6.9, p < 0.001]$; and, also it was also greater than that computed at note 19 $[\beta(34.7), t(4260) = 6.9, p < 0.001]$ (see **Table 5.5**, rows 12-14). The variability of this effect among participants was small ($VPC = 1.3$ %). *Post-hoc* comparisons between same pairs of notes in the two different conditions (i.e., UVL and UVF) demonstrate that the effect of leadership instruction is not mostly seen on a specific note of those selected for the analysis. For this reason, the analysis was not conducted at a deeper level (stage 3), i.e. on a specific significant note.

In summary, these results demonstrate that when the upper voice was instructed to lead, the designated leader showed overall a tendency to precede the designated follower; whilst, when the upper voice was instructed to follow, nobody tended to precede/lag the co-performer. This effect was consistent among participants, and the amount of leadership changed according to the time categories and the selected notes, but was not associated with any specific note of those selected for the analysis.

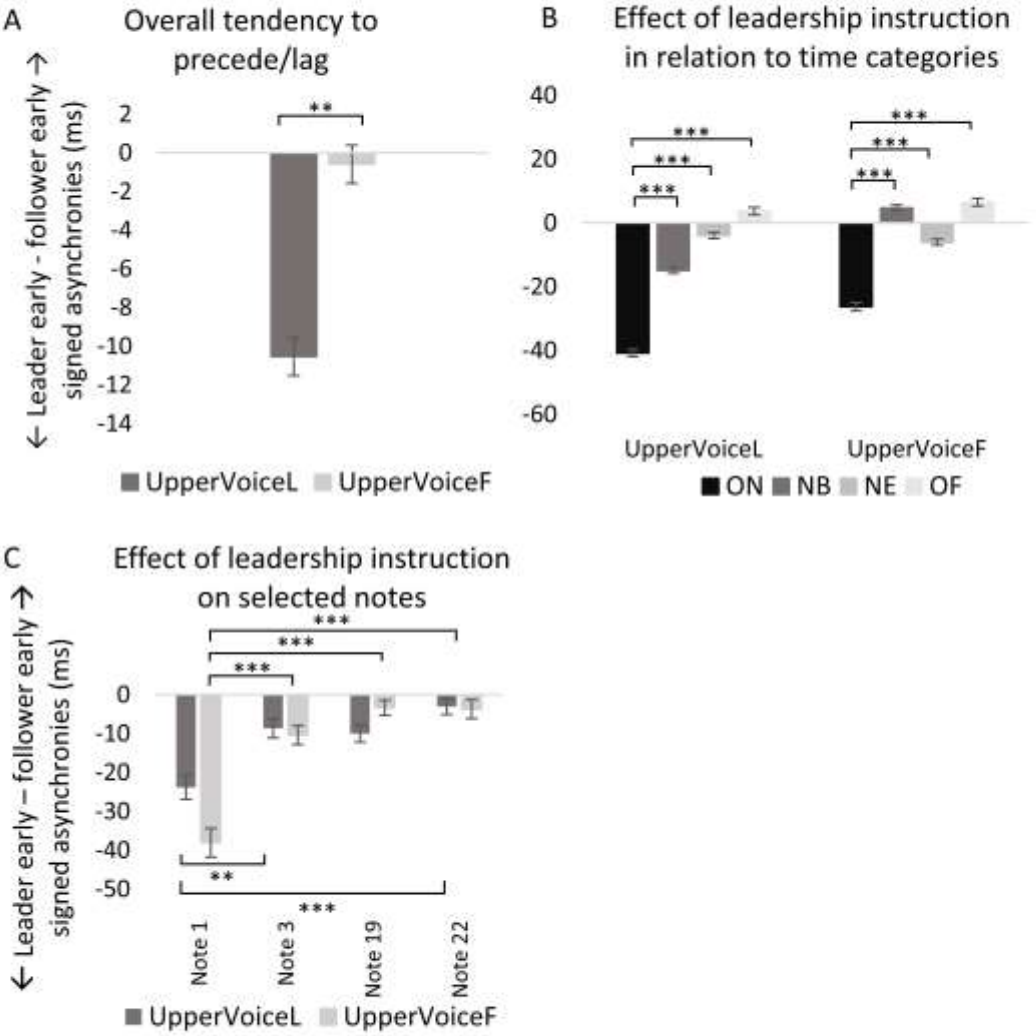


Figure 5.7 *Effect of leadership instruction on the tendency to precede/lag a co-performer for the overall performance (panel A); in relation to time categories (panel B); and, in relation to selected notes (panel C). Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $** \; p < 0.01$; $*** \; p < 0.001$.

Table 5.5 *Overview of the multilevel linear models developed to investigate the tendency to lead/lag, with the primary effects of leadership instruction, nested effects of crucial notes and time category, and the random effects of participants and chosen notes.*

Stage of analysis	Fixed effects	β coefficients and significance			Random effects	Row number
Stage1: Overall	Leadership	$\beta(-14.3)\ {*}{*}, t(25112) = -3.0$			Participants	1
	Time category nested	UVL	NB	$\beta(26.2)\ {*}{*}{*}, t(1247) = 6.6$	Chosen notes	2
			NE	$\beta(37.3)\ {*}{*}{*}, t(1996) = 9.5$		3
			OF	$\beta(43.9)\ {*}{*}{*}, t(1286) = 9$		4
		UVF	NB	$\beta(31.5)\ {*}{*}{*}, t(1271) = 7.9$		5
			NE	$\beta(20.8)\ {*}{*}{*}, t(2044) = 5.3$		6
			OF	$\beta(32.4)\ {*}{*}{*}, t(1327) = 6.6$		7
Stage 2: Notes subset	Leadership	$\beta(14.4)\ {*}{*}, t(4270) = 2.9$			Participants	8
	Crucial notes nested	UVL	3	$\beta(15.1)\ {*}{*}, t(4260) = 3.0$		9
			19	n.s.		10
			22	$\beta(20.7)\ {*}{*}{*}, t(4260) = 4.2$		11
		UVF	3	$\beta(27.7)\ {*}{*}{*}, t(4260) = 5.5$		12
			19	$\beta(34.7)\ {*}{*}{*}, t(4260) = 6.9$		13
			22	$\beta(34.4)\ {*}{*}{*}, t(4260) = 6.9$		14

n.s. $=$ not significant; ${*}{*}{*}\ p < 0.001.$ β - fixed effect coefficients on the predictor being considered - are given above with reference to the specified base level of the factor, i.e. NB, NE or OF versus the base level ON, and note 3, 19 or 22 versus the base level note 1.

5.4 Discussion

The study presented in this chapter investigated whether visual contact and assigned leadership roles contribute to interpersonal synchronization during singing duo performances. Three measures of interpersonal synchronization were considered: precision of synchronization as quantified by absolute asynchronies, consistency represented by SD of absolute asynchronies, and the tendency to precede or lag a co-performer indicated by signed asynchronies.

The presence or absence of visual contact between singers had a significant effect on the precision and consistency of synchronization, being better when the visual contact between singers was present, compared with when the visual contact was absent. In comparison, the results found in the pilot study with only two singing duos, which were reported in the previous chapter, indicated an increase in the consistency and precision of synchronization when there was no visual contact between singers in the case of duo 1 and duo 2, respectively. These apparently different results can be understood in light of the different sample sizes. Furthermore, visual contact, as observed in this chapter, also had an effect on the tendency to precede or lag a co-performer: without visual contact, the designated leader showed overall a stronger tendency to precede the designated follower, than in the presence of visual contact. These effects were consistent across participants and notes. These results expand on previous research focused on the effect of visual contact on instrumental ensembles which suggests that visual contact might affect synchronization during instrumental performances when auditory feedback is limited (Goebl and Palmer 2009) and in the presence of tempo changes (Kawase 2014; Bishop and Goebl 2015). The present study shows that visual contact might affect interpersonal synchronization also during singing duo performances. In addition, this study builds on previous investigations analysing interpersonal synchronization during ensemble performances, in which the tempo was controlled by a metronome, and musicians were clearly required to focus on timekeeping (Keller and Appel 2010; Goebl and Palmer 2009). Unlike these previous studies, the current work contributes to knowledge of the role of visual contact in interpersonal synchronization as emerging spontaneously during repeated performances rather than through enforce synchronization to some metronomic force.

The results also demonstrate that effects of visual contact on aspects of synchronization were seen most strongly at the onset of the first note. Precision and consistency observed at the onset of note 1 were better with visual contact, compared to when visual contact was absent. The tendency to precede the co-performer at the onset of note 1 was stronger when visual contact was absent than when the singers could see each other. These results show that visual contact might affect the synchronization of the onset of the first note, but musicians are able to compensate soon after, achieving a tighter interpersonal coordination, which also suggests optimal feedback adaptation. These findings support the pilot study reported in the previous chapter, which found that visual contact affected synchronization temporally computed at note beginnings, but not at note endings. These findings suggest the benefits of identifying strategies to improve ensemble music performance, refining rehearsal techniques and improving the experience of ensemble singing.

When performed with visual contact, precision and consistency computed at the beginning of phonation of the piece were different to those computed at the onset of note 19 (another simultaneous entry point) and beginning of note 22. These differences were amplified when performed without visual contact. The tendency to lead/lag at the beginning of phonation of the piece was not different from that computed at other onsets of the piece (i.e., note 19) or other note beginnings (i.e., note 3 and 22) with visual contact. However, without visual contact, the amount of leading at the onset of note 1 was greater than that computed at other onsets (i.e., note 19) and note beginnings (i.e., note 3 and 22). These results suggest that synchronization computed at the onset of note 1 might be different than other onsets of the piece, and note beginnings within a legato phrase. These differences might be intensified when visual contact is absent.

The researcher's instruction to act as leader or follower of the performance had no overall effect on the precision and consistency of synchronization, but leadership instruction had an effect on the tendency to precede/lag a co-performer. When the upper voice was instructed to lead, the designated leader tended to precede the follower by a small, but significant amount. Notably, when the upper voice was the designated follower, there was no clear separation of roles. These findings are consistent across participants and notes. These results complement the findings reported by Goebl and Palmer (2009) for piano duets performing piano duos in which the note ratio between the pianist is manipulated, and by Zamm, Pfordresher, and Palmer (2015) analysing piano duets performing the same part in unison and round. Overall, the results suggest that the effect of the instruction to lead or follow might depend on the piece being performed. The designated leader is more likely: i) to precede the performance of onsets if auditory feedback is limited when performing more notes than the designated follower (Goebl and Palmer 2009); ii) to lag the performance of the onsets when participants perform the same parts in a round (Zamm, Pfordresher, and Palmer 2015); and, iii) to precede a designated follower when the designated leader was performing the upper voice a two-part piece with a less clear separation of roles induced by the score, as found in this study. In addition, the current investigation shows that when the upper voice was instructed to follow, anybody tended to precede/lag a co-performer. This last finding suggests that trained musicians might have developed a compensatory behaviour, enabling them to maintain a tight and consistent synchronization, when the upper voice was instructed to follow.

Precision at ON was significantly larger compared with NB, when the upper voice was instructed to lead and also when instructed to follow, suggesting that precision at the beginning of phonation is larger than that at NB. Instructing the upper voice to lead appears to have intensified the difference in the precision of synchronization between ON and NB. The tendency to lead/lag was different based on the time category considered in relation to the leadership instruction, suggesting a bidirectional adaptation rather than a clear adaptation of roles. This finding corroborates the pilot study suggesting reciprocal adaptations between musicians that are not limited to the attack of the note, but are also associated with note beginnings, endings and offsets.

The synchronization characteristics reported in the current study (i.e., average absolute asynchronies of 58.99 ms; average SD asynchronies of 67.07 ms; median signed asynchronies of -4.06 ms) are in line with the literature analysing synchronization during naturalistic ensemble performance and controlled experiments, reporting typical mean absolute and SD asynchronies falling between 30 and 50 ms, and mean signed asynchronies often close to 0 ms (Keller 2014; Goebl and Palmer 2009; Timmers et al. 2014; Bishop and Goebl 2015; Zamm, Pfordresher, and Palmer 2015). Slight differences can be understood in light of the music material and/or the tempo used, and/or the techniques used to collect the data and extract onsets timing. In addition, the main characteristics reported in the current study were based on the measurement of synchronization temporally computed at ON, OF, NB, and NE, whilst the literature has mostly analysed synchronization at ON of each note, without differentiating between ON and NB, nor including NB or NE.

5.5 Limitations and Future Work

This study focused on the instruction to act as leader or follower, whilst interpersonal coordination patterns in music ensembles might be different if no leader/follower instructions are given, as debated in section *2.2.2.* This aspect is investigated in the following chapter by analysing interpersonal synchronization and the leader-follower relationships that emerge spontaneously in an advanced singing quintet, while no specific group roles are assigned.

It is important to consider the extent to which the effects of visual contact on the asynchronies in this study are perceived by listeners, especially in light of previous research showing that listeners might be sensitive to the degree of between-player asynchrony, when judging lack of togetherness in string quartet performances (Wing, Endo, Yates, et al. 2014). This is addressed in chapter *7*, which presents a study investigating the perception of synchronization in relation to the recordings collected and analysed here.

5.6 Conclusions

This chapter assessed the impact of visual contact and leader and follower relationships on synchronization between singers during singing duo performances. Results show that the presence and absence of visual contact between singers had a significant effect on the precision and consistency of interpersonal synchronization, and on the tendency to lead or lag a co-performer during singing duo performances. Precision and consistency were better when the singers could see each other than when they could not. The tendency to precede or lag a co-performer was greater without visual contact, and this effect was associated with the onset of note 1. These findings were consistent across performers. The instruction to act as the leader or follower of the performance did not affect the precision and consistency of interpersonal synchronization, nor the tendency to precede or lag a co-performer. The variability of these results among singers was small. Synchronization was found to change based on the time category considered, being often larger at the onset of phonation at the beginning of the piece. The absence of visual contact and instructing the upper voice to lead led to an amplification of these differences between time categories.

The next chapter goes on to investigate synchronization and leader-follower relationships emerging spontaneously across five rehearsals (i.e., without assigning specific roles), in relation to the leader-follower relationships and the rhythmical complexity of the pieces rehearsed. An additional analysis of tuning is conducted, and results are considered in light of the verbal discussions occurring during rehearsals, to further contextualise synchronisation within the wider scope of expressive ensemble performance.

6 Evolution of Synchronization across Rehearsals in a Singing Quintet

6.1 Introduction

The previous chapter described the effects of altered visual contact and the instruction to act as leader or follower on synchronization between singers during singing duo performances, through a repeated-measures design involving a single laboratory session. This design is typical of this field of research, as shown by the increasing number of studies into interpersonal synchronization investigating single performance sessions (Loehr and Palmer 2009; Goebl and Palmer 2009; Keller and Appel 2010; Glowinski et al. 2012; Badino et al. 2014; Kawase 2014; Zamm, Pfordresher, and Palmer 2015; Bishop and Goebl 2015; Timmers, Endo, and Wing 2013; Timmers et al. 2014). Even though this experimental design enables the analysis of the factors underpinning synchronization and leadership during a performance session, the evolution of synchronization and leader-follower relationships across rehearsals remains mostly unexplored, as noted in section *2.2.3*. The typically tight synchronization observed between members of Classical professional ensembles, reported in chapter *2* and confirmed in the previous chapter, may require practice over several rehearsals to be achieved. The investigation of the evolution of synchronization across rehearsals is, for this reason, of potential interest to music psychologists, to shed some light on the mechanisms leading to such a precise synchronization. This chapter[9] addresses this aspect through a longitudinal study following a newly formed, semi-professional singing ensemble from initial rehearsals to the performance stage, to investigate the evolution of synchronization across five rehearsal sessions during a four-month study period.

In addition, to investigate further the roles of leadership in singing ensembles, which were evaluated in the previous chapter by assigning specific roles of leader and follower in singing duos, this chapter observes the evolution of the leader-follower relationships emerging spontaneously during singing ensemble performances. As noted in chapter 2 (see section *2.2.2*), recent quantitative case studies assessing leadership in small ensembles revealed the relative leadership of the first Violin (Glowinski et al. 2012) and a complex pattern of leader-

---

[9] Most of this chapter, including figures and tables, is reported in D'Amario et al. (2018) and D'Amario, Daffern, and Bailes (2018a), and is © the authors, licensed CC-BY.

follower relationships rather than a more traditional division of roles (Timmers, Endo, and Wing 2013; Timmers et al. 2014); and, they suggest a link between leadership and the sharing of knowledge between co-performers (Badino et al. 2014). Moreover, qualitative studies focused on a singing octet and an 11-men singing ensemble suggest that leader-follower relationships might be democratically distributed in singing ensembles (Lim 2014; Page-Shipp, Joseph, and van Niekerk 2018). However, the leadership emerging spontaneously between musicians (i.e., without a designated leader/follower) and conceptualized as a tendency to precede/lag a co-performer during ensemble performances, has not been fully investigated in singing ensembles, nor in relation to its evolution over an extended period.

Furthermore, as noted in chapter 2 (see section *2.2.4*), previous studies conducted with piano duos suggest that interpersonal synchronization may also be affected by the different note ratios between the two parts when the auditory feedback of the upper part is limited (Goebl and Palmer 2009), the presence/absence of quavers within crotchet beats (Loehr and Palmer 2009), and the rhythmical characteristics of the piece, such as the non-simultaneous entrance of the two musical parts (Zamm, Pfordresher, and Palmer 2015). However, the study across rehearsals of the synchronization patterns associated with the different music characteristics of the piece currently lacks full investigation. This chapter addresses this aspect, observing the evolution of synchronization in relation to the rhythmical complexities of the pieces rehearsed.

Moreover, the analysis of interpersonal synchronization has been mostly conducted investigating duo performances (Goebl and Palmer 2009; Palmer et al. 2013; Keller and Appel 2010; Zamm, Pfordresher, and Palmer 2015; Bishop and Goebl 2017; Bishop and Goebl 2015; D'Amario, Daffern, and Bailes 2018b; D'Amario, Daffern, and Bailes 2018c) and quartet performances (Timmers, Endo, and Wing 2013; Timmers et al. 2014; Badino et al. 2014; Glowinski et al. 2012). The analysis of synchronization in larger ensembles lacks thorough investigation. This chapter centres on a singing quintet. Such an ensemble is common in the Western Classical tradition. The analysis remains feasible in a group of this size, but synchronization and leadership might change in this much more complex context, compared with the results reported in relation to duo and quartet performances (see section *2.1.1*).

Furthermore, as explained in section *2.3*, although it has been argued that interpersonal synchronization between musicians in music ensembles is one of the fundamental musical elements that contributes to the expressiveness of ensemble performances (alongside tuning, articulation, intensity, and timbre), research has mostly focused on the study of single parameters. During an expressive performance, synchronization, intensity, timbre, articulation, and tuning commonly co-vary deliberately to convey information about musical structure and expressive intentions to co-performers and audience (Keller 2014). The importance of these inter-individual co-variations imply the relevance of contextualising any finding related to synchronization, within the characteristics of the other parameters, advancing our understanding of synchronization within the wide scope of expressive performance. In this chapter, the study of the developmental aspects of synchronization in ensemble singing is explored in relation to the tuning systems adopted by the ensemble, to contextualise synchronization outcomes in light of tuning behaviours during expressive ensemble performances.

The focus of this chapter on tuning, in parallel with synchronization, reflects the crucial role that tuning plays in *a cappella* singing ensembles. Tuning represents a fundamental element in ensemble singing that boosts expressiveness in ensemble performance, at the forefront of critical reviews, director's manual and vocal coaches (Hansen 1964). Whilst intonation in the context of piano duos, the focus of several studies on synchronization (Goebl and Palmer 2009; Loehr and Palmer 2009; Keller and Appel 2010; Kawase 2014; Bishop and Goebl 2015), is not relevant to the study of expressive performance, tuning during *a cappella* performances is intrinsically dependent on the singers. During expressive ensemble performances, there are decisions, conscious or otherwise, that have to be made between the co-performers. The tuning system adopted, for example, represents an important expressive performance parameter. As reviewed in section *2.3* and in more detail in D'Amario et al. (2018), research analysing intonation in ensemble singing provides often sporadic and contrasting results regarding the tuning systems used. One research team, investigating tuning in four three-part singing ensembles, reported examples of both equal temperament (in which the octave is divided into 12 semitones of equal size) and just intonation (based on acoustically pure intervals, resulting from the overtone series) in the horizontal analysis of tuning and in the tuning of major and minor third intervals (Devaney, Mandel, and Fujinaga 2012). Howard (2007b) observed pitch drift towards just intonation in *a cappella* singing quartets, whilst Howard (2007a) reported pitch drifted beyond the just intonation prediction and a long way far from equal temperament in a different *a cappella* singing quartet performing one of the pieces used in the previous study. This chapter aims to shed some light on the evolution of the tuning adopted across rehearsals in relation to specific tuning systems.

Finally, in addition to the parallel analysis of tuning and note synchronization, analysis of the verbal discussions between singers during the five rehearsal sessions is also conducted. Verbal discussions are considered to investigate whether and, eventually, how aspects of tuning and synchronization are deliberately approached during rehearsal.

This novel mixed-method repeated measures design spans several rehearsals to analyse the evolution of synchronization during a four-month study period, combining both quantitative performance data (i.e., synchronization and tuning data) with observational frameworks. A number of research questions have been formulated:

- Do interpersonal synchronization and/or the leader-follower relationships between singers change with practice in a singing quintet?
    - Do these changes, if any, differ in relation to the contrasting musical features of the pieces rehearsed?
- Does the singing quintet produce a pitch drift representative of just intonation predictions or maintain horizontal tuning in equal temperament?
    - Do these horizontal tuning trends change pre- and post-rehearsal?
    - Do these horizontal tuning trends change longitudinally over rehearsal sessions spanning four months?
- Does the singing quintet tune thirds within chords towards just intonation or equal temperament?
    - Do these vertical tuning trends change pre- and post-rehearsal?
    - Do these vertical tuning trends change longitudinally over rehearsal sessions spanning four months?
- How do singers approach synchronization and tuning in rehearsal?
- Which rehearsal strategies are associated with the synchronization and tuning outcomes of this ensemble?

Although this study is explorative in nature, it was hypothesized that rehearsals increase synchronization between singers during singing quintet performances, and that this effect depends on the rhythmical complexity of the piece being performed. It was also conjectured that this quintet is characterized by mutual adaptations of leader-follower roles, as found in quantitative analyses of string quartet performances (Timmers, Endo, and Wing 2013; Timmers et al. 2014; Glowinski et al. 2012), and qualitative studies of a singing octet (Lim 2014) and an 11-men singing ensemble (Page-Shipp, Joseph, and van Niekerk 2018). Moreover, it was hypothesised that tuning becomes more consistent across rehearsals. Finally, it was conjectured that the singers' focus, if any, on aspects of synchronization and tuning during rehearsal in order to deliver an expressive performance, decreases across the study period, as found in studies analysing verbal discussions during ensemble rehearsals, showing a shift from basic tasks in early rehearsals, including tuning and timing, to more interpretative tasks in late rehearsals (Ginsborg, Chaffin, and Nicholson 2006).

6.2 Method

6.2.1 Participants

Ethical approval for the study (with reference D'Amario070817) was obtained from the Physical Sciences Ethics Committee (PSEC) at the University of York (UK). A soprano, mezzo, mezzo, tenor, and bass singing quintet took part in the study (three females, age $Mdn = 23\ yo,\ Range = 6\ yo$). Singers were Master of Arts students in ensemble singing at the Department of Music of the University of York. At the time of the study, the quintet was a newly formed ensemble established as a regular quintet working towards performances and Masters exams. They had met for only one rehearsal prior to the beginning of the first session, but rehearsed regularly throughout the duration of the study in preparation for their final exam. All singers had formal singing training with a professional singing teacher ($Mdn = 8\ y, Range = 13\ y$), and extensive experience performing in choirs ($Mdn = 10.8\ y, Range = 11\ y$) and in singing ensembles such as duos, trios and quartets ($Mdn = 7\ y, Range = 8\ y$). In addition, the bass had 12 years of experience conducting, and five years composing. They reported not having absolute pitch, and having normal hearing, except for the bass who reported sporadic tinnitus during his life that did not affect his singing during the rehearsal sessions.

6.2.2 Stimuli

This investigation made use of two chorales composed by Johann Sebastian Bach: one piece was the chorale *Jesu, mein Hort und Erretter* from the Cantata BWV 154 *Mein liebster Jesus ist verloren*; the other was the chorale *Nun danket alle Gott* from the Cantata BWV 192. The pieces were arranged for the singing quintet ensemble in the study by the author. These chorales were chosen for their structural characteristics: two short pieces, mostly homophonic, with different melodic contours and harmonic structure from each other, and feasible for mastery within five, short rehearsal sessions. The two arranged pieces comprise eight phrases, to be sung legato to the vowel */i/*. This vowel was chosen by the singers that took part in the pilot study described in chapter *4*, investigating synchronization in singing duo performances. The same vowel was used in the present study for consistency with the previous investigations.

To facilitate the analysis of synchronization and tuning based on the fundamental frequency (f_0) tracking described in section *6.2.6*, the two pieces were arranged based on the following criteria: i) avoiding repeated notes, and ii) limiting semitones. The detection of note beginnings (NBs) and note endings (NEs) within a legato phrase and, consequently, the analysis of interpersonal synchronization and tuning can be potentially difficult to compute from the f_0 track of audio recordings when melodies move chromatically, since the expected vibrato range for classical singers might span a semitone, as argued in the pilot study and in the main study described in the previous chapter. Similarly, true NBs of repeated notes can be difficult to detect from the f_o estimates, if singers do not produce a noticeable pause in phonation between notes, as highlighted in chapter 2 (see section *2.5.3*). Two pieces with these characteristics (i.e., without repeated notes, and only few semitones), which maximize asynchrony detection and tuning analysis were difficult to find, so arrangement of the pieces was preferred. As a result of the arrangement process, the resulting pieces differ to the original chorales in several ways. For example, the melody of the bass in the first two bars of piece A remains very similar to the original, whilst that of the soprano has been changed to avoid the repeated notes of the first bar and the semitone of the second bar present in the original chorale, which would have impacted the performance of the TIMEX algorithm.

The pieces do not include any expressive markings and singers were invited to develop their own expressive interpretation, in order to investigate aspects of interpersonal coordination, including temporal coordination and tuning, that might emerge spontaneously in support of expressive performances. No clear leader-follower roles are defined in the pieces to allow relationships between singers to emerge during rehearsals. The two pieces varied in their rhythmic, melodic and harmonic structure, as follows:

- Piece A: characterized by a clear homophonic structure, with a constant 1:1 note ratio across performers (i.e., equal note length for each voice), featuring simultaneous entries and breaths, and a stable rhythm, as shown in **Figure 6.1**.
- Piece B: characterized by a different harmonic structure, and a different and more complex melodic and rhythmic structure, including ornamentations within each part that varied the note ratio across performers. Entries were systematically manipulated in a way that, except for the first simultaneous entry, each singer had one occasion to start the phrase on an up-beat ahead of the others (i.e., the mezzo 2 entered ahead of the others in bar 3, the tenor entered earlier in bar 5, the bass in bar 7, the soprano in bar 9, and the mezzo 1 in bar 11), as shown in **Figure 6.2**.

The contrasting characteristics identified above were introduced to investigate whether the evolution of synchronization might vary depending on the properties of the piece being sung. Both pieces were used for the analysis of temporal coordination, and only piece A was used for the investigation of tuning, because of its homophonic nature featuring chords with no passing notes. Chords that do not include passing notes may be more exposed to pitch drift, as they have more time to adapt, and therefore they are of particular interest to the study of the tuning systems adopted across rehearsals.

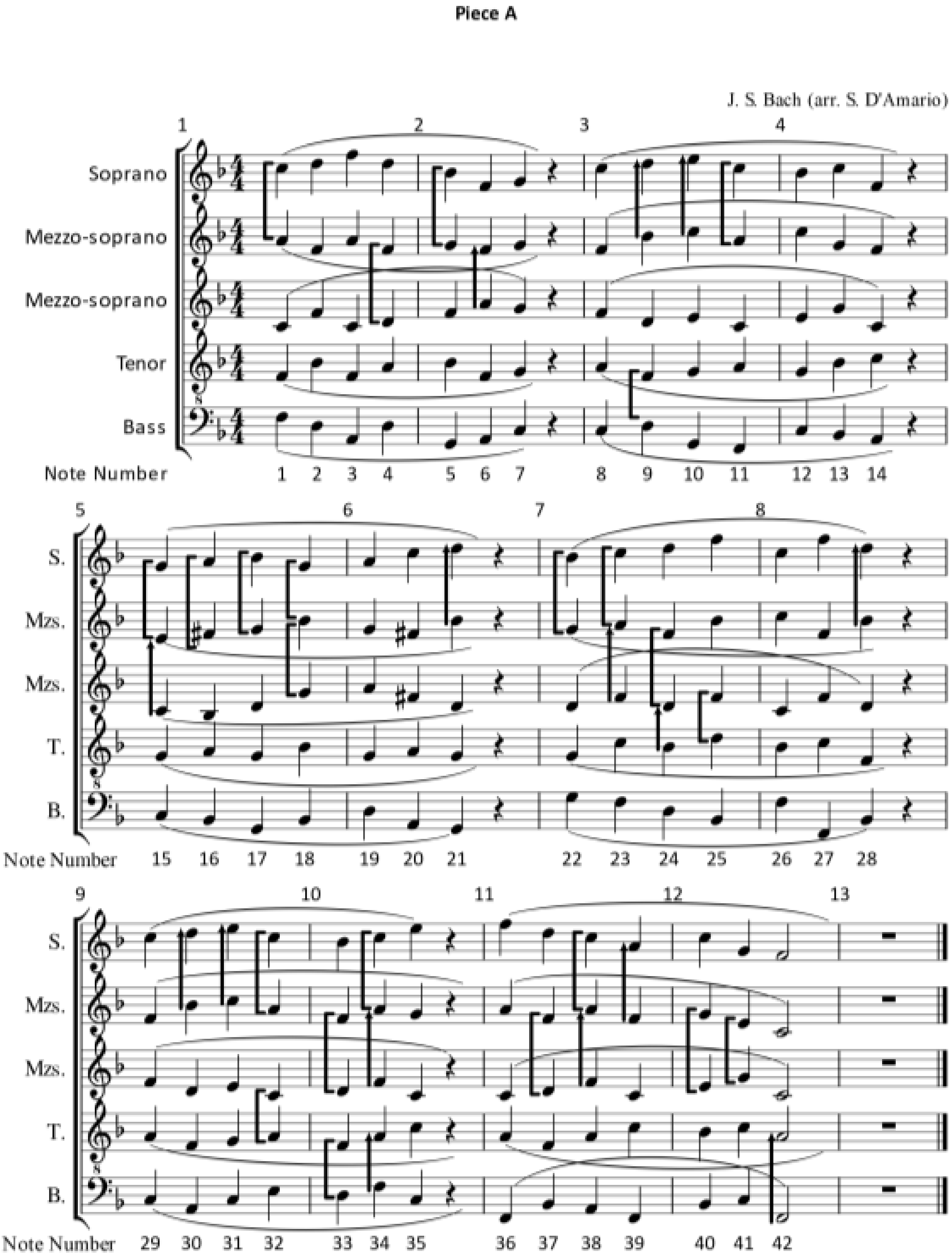


Figure 6.1 *Piece A used for the study. The full set of notes was used for the analysis of synchronization between singers and horizontal tuning. The figure displays the major and minor thirds, highlighted with arrows and brackets respectively, that were selected for the analysis of vertical tuning.*

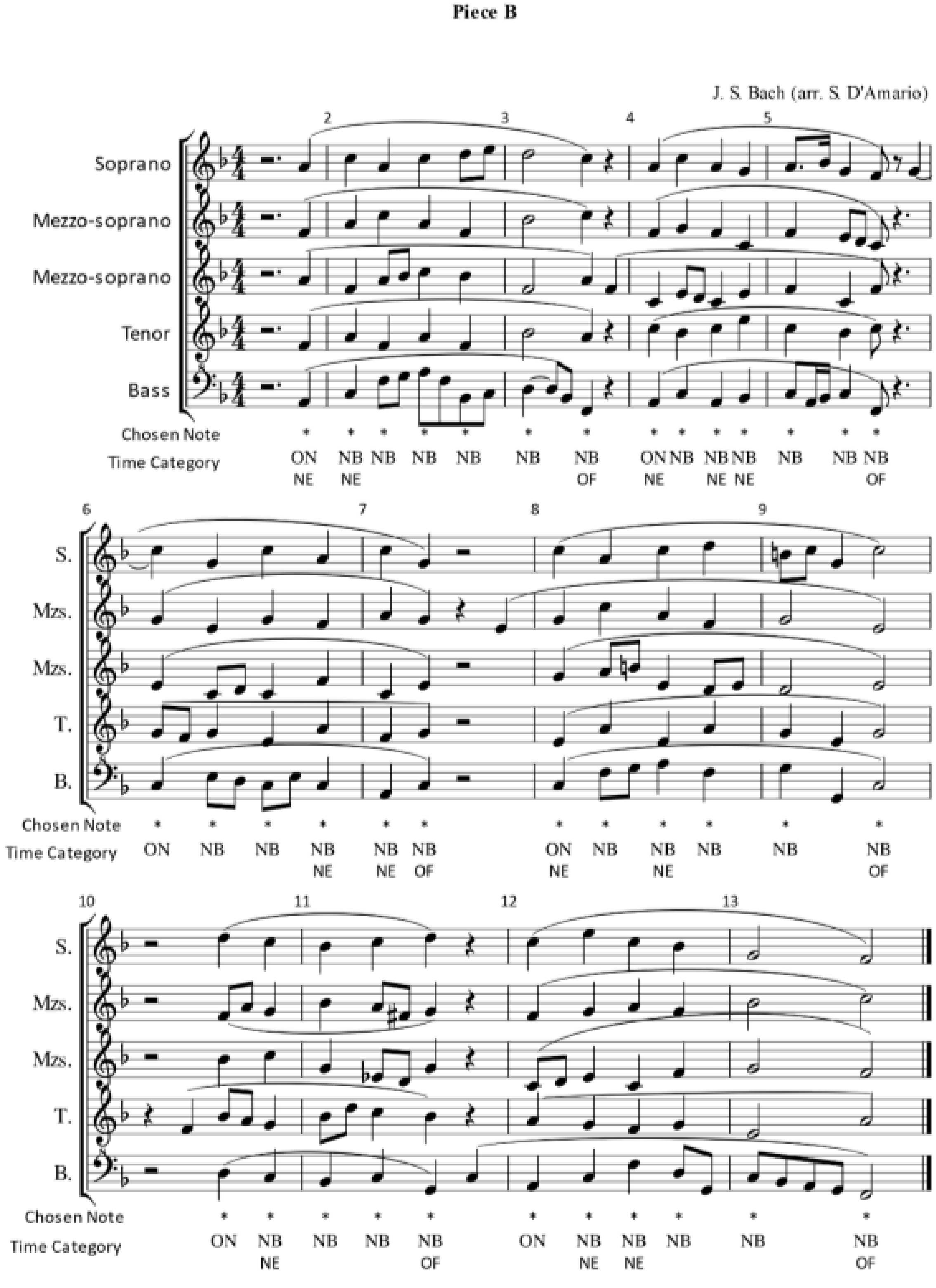



Figure 6.2 *Piece B used in the study for the analysis of synchronization, showing the notes, highlighted with *, and the time categories (onset, ON; offset, OF; note beginning, NB; and note ending, NE) upon which the analysis is based.*

6.2.3 Design

This investigation is a longitudinal study of five rehearsal sessions based in a laboratory setting. During each rehearsal session, the above pieces were practised, and three repeated performances of the pieces were recorded pre- and post-rehearsal. This study resulted in a 5 [rehearsal sessions (R)] × 2 (pieces) × 2 [stages of rehearsal, 1 pre- (pre) and 1 post-rehearsal (post)] × 3 [repeated performances for each stage of rehearsal (RP)] design, featuring a total of 30 repeated performances per piece, recorded during the course of the study. The order of recording and rehearsing the two pieces was randomized within rehearsals. Therefore, in rehearsal 1, 4 and 5, singers first recorded-rehearsed-recorded piece A, then piece B; but, in rehearsal 2 and 3, the quintet recorded-rehearsed-recorded piece B first, followed by piece A.

Additionally, a short questionnaire was administered at the end of the last rehearsal session, and singers were interviewed a month after the end of the experiment as part of the PhD study into the verbal interactions of the ensemble by Nicola Pennill; the detailed results of the questionnaire and interviews will be thoroughly reported in her PhD thesis.

6.2.4 Apparatus

The experiment took place in a bespoke recording studio of the Department of Electronic Engineering at the University of York. The room was $5.2\ m \times 7.6\ m$, the ambient noise level was 37 dB(A), and the RT60 reverberation time was 0.32 s. Therefore, the room was an environment not alien to the singers in terms of acoustic (relatively dead, typical of a practice room or recording studio) and look (a recording studio with acoustic panelling on the walls and recording equipment). The quintet stood in a semi-circle of approximately 1.5 m in radius in the sequence soprano (S1), mezzo 1 (S2), mezzo 2 (S3), tenor (S4) and bass (S5), with S1 opposite S5. This ensemble was not a choir with several voices per part, but a vocal ensemble. This aspect limits the relevance of any comparisons to spacing in choirs. Nevertheless, some considerations have been made regarding the placement of voices. The 1.5 m radius was chosen so the distance between adjacent singers was $2\ \pi \times \frac{1.5}{2x4} \approx 1.2\ m$. This is larger than a conventional choral spacing, but close to what is often considered preferable on a concert podium (Daugherty 1999). In combination with the fact that each singer was alone per part, this distance of 1.2 m indicates that the self-to-other ratio was high, and that each singer could very easily hear their own voice above the others.

As in the pilot study and in the main study reported in the previous two chapters, the following sets of data were acquired: acoustic data, based on a stereo condenser microphone (Rode NT4) providing right and left outputs, and on head-mounted close proximity microphones (DPA 4065); and, electrolaryngograph (Lx) recordings, using Lx electrodes from Laryngograph Ltd. (www.laryngograph.com). Stereo recordings were collected and used for the perception study reported in chapter *7*, and were not used to measure interpersonal synchronization. The stereo microphone was a twin cardioid microphone pointing towards S3, with its main axes pointing at S2 and S4; and, it was placed at equal distance in front of the singers at approximately 1.5 m from the lips. This stereo microphone was chosen as being most sensitive to the area in front of the microphone capsule, where the singers stood, while picking up minimal and marginal noise from the rear and sides, respectively.

Similarly to the two studies reported in the previous two chapters, Lx electrodes were chosen as they allow f_o evaluation of the individual voices, and were placed on either side of the neck at the level of the larynx. Closed proximity microphones were used and acoustic recordings collected to analyse the synchronization in the very few cases, as tested in chapter 3 (see section *3.3.2*), in which the Lx signal can be unusable, as demonstrated during the test of the reliability of the Lx signal. Each head-mounted microphone was placed on the cheek of the singer at approximately 2.5 cm from the lips.

Each Lx box was connected to a preamplifier (ART CleanBox Pro) to reduce noise and interference between the Lx boxes over long cable runs. The 12 outputs (five Lx with preamplifiers, five head-mounted microphones, and the stereo microphone comprising right and left channel) were attached to a multi-channel audio interface (Focusrite Liquid Saffire 56), which was connected to a PC; the 12 outputs were then recorded using a digital audio workstation (Reaper 5.40) at a sampling frequency of 44.1 kHz and 24-bit depth.

Rehearsals were also video-recorded with a tripod-mounted video camera (Sony MV1 Music Video recorder), with a unidirectional 120 degree XY stereo microphone. This was done for the analysis of verbal interactions as part of the PhD thesis by Nicola Pennill.

6.2.5 Procedure

The singing quintet was invited to five rehearsal sessions over a four-month period, from September 2017 to January 2018. In each session, the singers rehearsed each piece for 10 minutes, and performed the pieces three times before and after each rehearsal, endeavouring to make each repetition an individual performance. Each session was approximately one-hour long. Singers were fully engaged during all rehearsals. Prior to the first rehearsal, participants were asked to fill in a background questionnaire and consent form.

The first four rehearsal sessions were approximately three weeks apart from each other, as shown in **Table 6.1**. The fifth lab session was originally planned three weeks after the fourth session, which was two days before the singers' Masters exam. Due to illness and the Christmas break, the exam was postponed until eight weeks after the fourth rehearsal, and the fifth lab session took place two days before the ensemble members' formal performance exam, set up in the form of a public concert. This was designed to conclude the analysis of synchronization and tuning at a time when the ensemble should be at its most cohesive, after four months practice at a performance standard. Singers were not aware of the purpose of the study.

Table 6.1 *Rehearsal sessions across a 16-week period*

Rehearsal number	1	2	3	4	5
Week	1	3	5	8	16

Singers were required to work on expressiveness, working towards a final performance of the stimulus pieces at the end of the term of study. They were left free to use the short rehearsal sessions in any way they chose to create an expressive interpretation. This was designed to encourage a realistic approach to rehearsal, promote a development of the quintet (although the two pieces rehearsed were not performed on stage at the end of the study term), and contextualise the results within the broader scope of expressive ensemble performances. Singers received the score of the stimuli on the day of each lab session to practise and perform the piece, but the author retained the score at the end of each sessions, so singers were not able to rehearse the pieces between lab sessions. This allowed the author to record and analyse any changes in the development of synchronization or tuning relating to the given pieces during the term of study. Singers performed and rehearsed the pieces *a* *cappella*, and an audience was never present. A reference pitch A3 was given on a diapason before the three repeated performances recorded pre- and post-rehearsal. The quintet was free to set their own tempo.

6.2.6 Analysis

The following three sections provide details of the three parallel analyses that were implemented in this study, i.e. analysis of synchronization, tuning, and verbal discussions.

For both synchronization and tuning analysis, Lx and audio recordings from the head-mounted microphones, were imported in Praat (Boersma 2001; Boersma and Weenink 2013) as .wav files, and two sets of data were extracted from each recording: the f_o estimates in Hertz and the corresponding timestamps in milliseconds with a time step of 1 ms, as in the previous studies described in chapter *4* and chapter *5*. These two data sets were then entered into Excel as a tabular list of data. The TIMEX algorithm described in chapter *3* and implemented in chapter *4* and chapter *5*, was used in the current study to extract the four time categories described in section *3.2*, i.e. onsets (ON) and offsets (OF) of phonation, and note beginnings (NB) and note endings (NE) within the sung legato phrases. As noted in the previous chapters, the algorithm relies mostly on the Lx recordings, but automatically scrutinizes the acoustics data from the head-mounted microphones when the Lx signal is too weak to be detected. The acoustic recordings were also scrutinized in cases of soft phonation, since the Lx signal may not reveal low amplitude vocal fold vibrations (Sten Ternström and Johan Sundberg, personal communication, June 2017).

6.2.6.1 Synchronization

As in the previous two chapters, the data extraction automated through TIMEX was visually cross-validated by the author. Note errors due to the singers performing the wrong notes (i.e., entering or delaying the notes for more than 50 % of its expected values) were less than 0.05 %, identified comparing Lx and audio recordings with the notated scores. These errors were excluded from the analysis of synchronization. In addition, manual detection of NBs and NEs was necessary when the algorithm failed to detect timing when the melody moved chromatically; in these cases, manual changes were less than 2 % of the data set. Finally, manual changes were necessary for an additional 42 % of the data set, in cases in which the singers performed a vibrato (within the same note) with an extent larger than a semitone; in these cases, the algorithm identified the frequency oscillations of the vibrato as note beginnings/endings, failing to detect the vibrato. The overall manual changes related to errors associated with timing, semitones, and vibrato were around 44 %.

All notes from piece A, the clearly homophonic piece, were selected for the investigation of interpersonal synchronization. Conversely, the analysis of temporal coordination in piece B was focused on notes being relevant to synchronization, as shown in **Figure 6.2**.

Similarly to the previous two chapters, asynchronies were then calculated for each pair of singers, subtracting the timing of one singer from that of another, such as soprano minus mezzo, and soprano minus tenor. This procedure gave a matrix with a total of 20 channels of asynchronies. These channels were implemented for each time category (i.e., ON, NB, NE, and OF) selected for the analysis of interpersonal synchronization.

Asynchronies that fell outside three times the interquartile range (IQR) were automatically detected as extreme outliers in SPSS (IBM SPSS Statistics v. 24) and excluded from the analysis. The identification of outliers was run for each time category, pre- and post-rehearsal condition, piece and rehearsal.

The same measures of synchronization were used as in the previous chapter. They were:

- <u>Precision</u> of synchronization, as quantified by absolute asynchronies, and expressed in milliseconds.
- <u>Consistency</u> of synchronization, as quantified by SD of absolute asynchronies expressed in milliseconds, and computed across the 20 channels (i.e., pairs of singers) for each time category.
- <u>Tendency to precede or lag</u> a co-performer, as quantified by signed asynchronies, expressed in milliseconds.

Multilevel linear models were then implemented step by step for each response variable (i.e., precision, consistency, and tendency to precede/lag), to investigate the fixed effects of rehearsal number and piece (the last one nested within rehearsals) as shown in **Table 6.4**. Notes, time categories, and stage of rehearsal (i.e., pre- and post-rehearsal) were entered as random effects in the models investigating precision, consistency and tendency to precede/lag. Pairs of singers were also entered as random effects in the models investigating the precision of synchronization and the tendency to precede/lag. As noted earlier in the thesis, multilevel linear models were chosen because they reinforce the statistical power of the analysis providing an assessment of the variability of the fixed effects across random effects (Gelman and Hill 2007). The models were implemented in R Studio (RStudio 2015), using the lme4 package. A Bonferroni correction was implemented for multiple multilevel linear models, dividing the critical value (0.05) by the number of comparisons being made, three, corresponding to the total number of models developed for the three response variables. For this reason, the alpha level was set at $p = 0.016$.

In addition, the tendency to precede/lag a co-performer was investigated analysing the temporal rank order across the five singers for each time category/note/repeated performance/rehearsal. As shown in **Figure 6.3**, the temporal rank order for note beginnings and onsets was analysed identifying the entrance temporal position of each singer from position 1 (singer who preceded all co-performers), to position 2 (singer who entered next), to position 3, 4, up to position 5 (singer who lagged all co-performers). Similarly, the sequence for note ending and offset was based on the analysis of the exit temporal position.

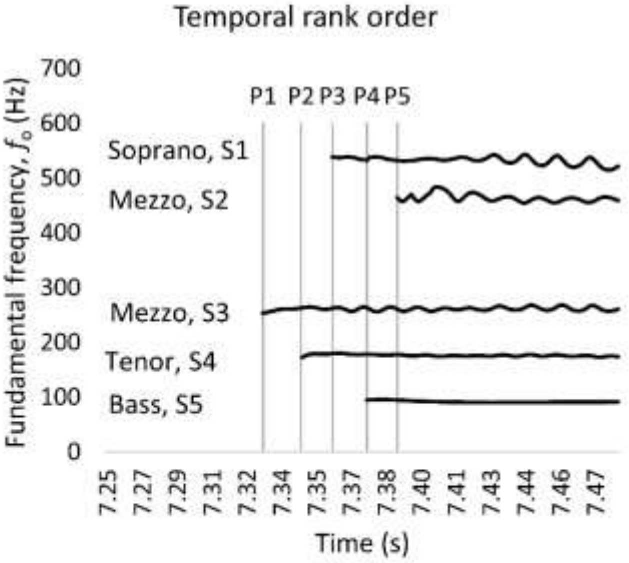


Figure 6.3 *Excerpt from the 5 Lx signals showing the temporal rank order regarding the entrances of the 5 singers performing the onset of the first note of piece A recorded during rehearsal 5. Based on the temporal entrance order from position 1 (P1) to position 5 (P5), the resulting temporal sequence observed here was S3-S4-S1-S5-S2, where S3 and S2 were the first and last singer to precede and lag all-co-performers, respectively.*

For each singer, the tendency to occupy a given position was measured counting the number of occurrences spent in that position, i.e. the observed frequency on a given position. This was computed for each position/rehearsal/singer, taking all notes, time categories and both pieces together. Then, three aspects of the effect of rehearsals on the tendency to precede/lag were analysed:

- The interaction between rehearsal (1-5) and position (1-5) for each singer, through Pearson's chi-square tests for independence. The test indicates whether the frequencies observed in each position are independent from the rehearsals. A total of 5 Pearson's chi-square tests were conducted, one for each singer.
- The interaction between rehearsals (1-5) and the frequencies at which each singer sang in position 1, through a Pearson's chi square test for independence. This analysis demonstrates whether there is a significant relationship between rehearsals and the number of occurrences each singer spent in position 1. One chi-square test was implemented, including the frequencies that each singer occurred in position 1 in each rehearsal.
- The distribution of the tendency to precede all the others computed across the five singers for each rehearsal. This investigation was conducted through a total of five chi-square goodness of fit tests (i.e., one for each rehearsal), comparing the observed frequency distribution with an equal frequency distribution. These analyses informed whether the tendency to precede all co-performers was equally distributed in each rehearsal. Whilst the Pearson's chi square test for independence presented in the previous point is an omnibus analysis that investigates the interaction between rehearsals and frequencies in position 1 across the five rehearsals, the five chi-square goodness of fit tests allows for a focused analysis on the distribution at each rehearsal.

A Bonferroni correction was carried out for the 11 chi-square tests implemented in this study. The alpha level was set at $p = 0.0045$, obtained by dividing the critical value (0.05) by the number of chi-square tests (i.e., $0.05/11$).

6.2.6.2 Tuning

As noted in section *6.2.2*, the analysis of tuning was based on Piece A, the clear homophonic piece, shown in **Figure 6.1**. Two aspects of tuning were analysed: i) horizontal tuning, and ii) vertical tuning, as shown in **Table 6.2**. After extracting beginnings and endings of each note through TIMEX, a macro was implemented to compute the mean frequency in Hertz of each note. The automated data computation was then visually cross-validated by the author, and notes at which pitch errors occurred, due to signal processing issues (i.e., weak signal) or the singers performing wrong notes (i.e., featuring a measured deviation from the expected ET value greater than 130 cents), were 1.9 % of the full data set. They were identified comparing the data values extracted with Lx and audio recordings, the notated scores, and the equal temperament (ET) predictions (see next paragraph for more detail); these values were excluded from the analysis.

Table 6.2 *Aspects and parameters of tuning investigated, and corresponding recordings and dataset.*

Aspect	**Parameter**	**Recordings**	**Dataset**
Horizontal deviation	Pitch drift, and tuning consistency and dispersion	Lx and audio	Deviation for each note/singer/repetition
Vertical deviation	Tuning stability, consistency and dispersion	Lx and audio	Deviation for major and minor thirds

In order to analyse the pitch drift during each of the performances, a reference set of f_o is required for the tuning systems of interest; in this case, ET (characterized by adjacent semitones of equal size) and just temperament (based on the acoustically pure intervals of the harmonic series), in line with previous investigations analysing tuning in *a cappella* SATB ensembles (Howard 2007a; Howard 2007b). These reference f_o were calculated as frequency multipliers to the tonic of the key of the chorale (see **Figure 6.1**), which is F as it is in F major, and the starting note of the tenor part (F3) was selected. The procedure for calculating the ET ratios involved multiplication (division) by the twelfth root of two to move up (down) by a semitone. The procedure for calculating the just ratios has two steps: (a) within a chord the intervals are calculated using integer harmonic ratios depending on the interval (e.g. a fifth is 3/2, a major third is 5/4, a minor third is 6/5 etc.), and (b) chord to chord where a search is carried out to find the nearest harmonic ratio between one of the notes of each of the chords in the following order: unison, octave, fifth, fourth, major third, minor third, etc. Further details on tuning systems and frequency ratios can be found in Howard and Angus (2017). The measured f_o values were entered into the spreadsheet, and the f_o of each sung note was divided by the measured f_o value for the first note (F3) of the tenor part which was the reference note for the analysis as indicated above. To establish how close the sung notes were to equal temperament or just temperament, the measured frequency ratios were divided by the equal (just) tempered ratios. For the analyses presented below, the results have been converted to cents (1 cent is one hundredth of a semitone) to enable comparisons to be made.

The horizontal analysis was based on the whole set of notes (i.e., 42 notes) included in the piece. A total of 15 major thirds and 23 minor thirds across parts were selected for the vertical analysis, as shown in **Figure 6.1**. The thirds were simple intervals, except for one compound major third, between bass and tenor in note 42, which was also selected for the analysis. This interval was considered relevant to the analysis of thirds, being the last chord of the piece. Three metrics of horizontal and vertical tuning were measured as follows:

- Pitch drift, as indexed by the pitch deviation from ET and just intonation.
- Tuning consistency, as indexed by the SD of measured deviations computed for each repeated performance, pooling the 42 notes or the selected thirds to analyse horizontal or vertical consistency, respectively.
- Tuning dispersion, as indexed by the range of measured deviation computed across notes or selected thirds for each repetition, similarly to the procedure implemented for tuning consistency analysis

Multilevel linear-models of the response variables (i.e., f_o deviation from predicted values, SD and range of measured deviation) were then implemented to test the primary fixed effects of rehearsal number, and the fixed effects of rehearsal stage (i.e., pre- and post-rehearsal) nested within rehearsal number, as shown in **Table 6.3**. Note, repeated performance, and singer number were also entered as random variables in the models investigating the horizontal tuning across all notes. Repeated performance number, interval and pair number were inputted as random variables in the models analysing the major and minor thirds.

A Bonferroni correction was implemented for multiple multilevel linear modelling, and the alpha level was set at $p = 0.0055$, based on a total of nine tests (see **Table 6.3**).

Table 6.3 *Multilevel linear models implemented in the study for the analysis of tuning*

Response variable	**Primary fixed effects**	**Nested fixed effects**	**Random effect**	**Data set**
Drift	Rehearsal number	Rehearsal stage	Note, singer and performance number	All notes
Consistency	Rehearsal number	Rehearsal stage	Note, singer and performance number	All notes
Dispersion	Rehearsal number	Rehearsal stage	Note, singer and performance number	All notes
Drift	Rehearsal number	Rehearsal stage	Performance and interval number, singer pair	Major thirds
Consistency	Rehearsal number	Rehearsal stage	Performance number	Major thirds
Dispersion	Rehearsal number	Rehearsal stage	Performance number	Major thirds
Drift	Rehearsal number	Rehearsal stage	Performance and interval number, singer pair	Minor thirds
Consistency	Rehearsal number	Rehearsal stage	Performance number	Minor thirds
Dispersion	Rehearsal number	Rehearsal stage	Performance number	Minor thirds

6.2.6.3 Verbal interactions

The verbal discussions between singers during the five rehearsals were transcribed by Nicola Pennill from the video recordings and the content scrutinised in relation to singers' specific reference to synchronization and tuning, and the results reported within this chapter. The time spent on debates regarding synchronization and tuning aspects was computed for each session, and the content of the verbal discussions analysed in relation to chord and bar number.

6.3 Results

6.3.1 Synchronization

This section presents the results of the analyses of the fixed effects of rehearsal and piece on interpersonal synchronization. The β - fixed effect coefficients - of rehearsal number (i.e., R1-R5) and piece (i.e., piece A and piece B) on the predictor being considered (i.e., precision, consistency, and tendency to precede/lag) are given below and in **Table 6.4** with reference to the specific base level of the factor (i.e., rehearsal 2, 3, 4 and 5 *versus* the base level rehearsal 1, and piece B *versus* piece A). The β coefficient indicates that for each 1 unit increase in the predictor being considered, the effect of the given predictor changes by the amount specified by the β coefficient.

6.3.1.1 Precision

Precision of synchronization was on average 62.59 ms $(SD = 95.46\,ms)$ computed by averaging the mean absolute asynchronies collapsing the entire data set, irrespective of rehearsal number, piece, note, and time category. Results from the multilevel linear modelling as explained in section *6.2.6* show that precision of synchronization improved from the first to the last rehearsals. As shown in **Figure 6.4A** and **Table 6.4**, compared to the baseline in rehearsal 1, precision improved in rehearsal 2 $[\beta = -9.7, t(40505) = -5.4, p < 0.001]$; in rehearsal 3 $[\beta = -6.7, t(40505) = -3.8, p < 0.001]$; and also in rehearsal 4 $[\beta = -8.8, t(40505) = -4.9, p < 0.001]$ and rehearsal 5 $[\beta = -11.9, t(40505) = -6.6, p < 0.001]$. Precision in the synchronization of piece A was better than that of piece B in all rehearsal sessions, as shown in **Table 6.4** and **Figure 6.4B**. The variance partition coefficient (VPC) among pairs of singers, notes, time categories and pre- and post-rehearsal was 0.009, 0.067, 0.030, and 0.00026. This indicates that only 0.9 %, 6.7 %, 3 % and 0.026 % of the variability of precision of synchronization over 5 rehearsals can be attributed to pairs of singers, note, time categories and pre- and post-rehearsal, respectively. As shown in **Figure 6.4B**, *post-hoc* comparisons, Holm corrected for multiple comparisons, revealed that precision in the synchronization of piece A improved in rehearsal 2 $(M = 49.2\,ms, SD = 43.4\,ms)$ compared with rehearsal 1 $(M = 58.7\,ms, SD = 50.1\,ms, t = 6.4, p < 0.001)$. Similarly, precision in piece B was better across rehearsals, as it improved in rehearsal 2 $(M = 80.0\,ms, SD = 140.1\,ms)$ compared with rehearsal 1 $(M = 106.9\,ms, SD = 50.1\,ms, t = 11.6, p < 0.001)$, in rehearsal 4 $(M = 71.8\,ms, SD = 116.3\,ms)$ compared with rehearsal 3 $(M = 81.5\,ms, SD = 130.1\,ms, t = 4.0, p < 0.001)$, and in rehearsal 5 $(M = 62.0\,ms, SD = 90.5\,ms)$ compared with rehearsal 4 $(t = 3.9, p < 0.01)$.

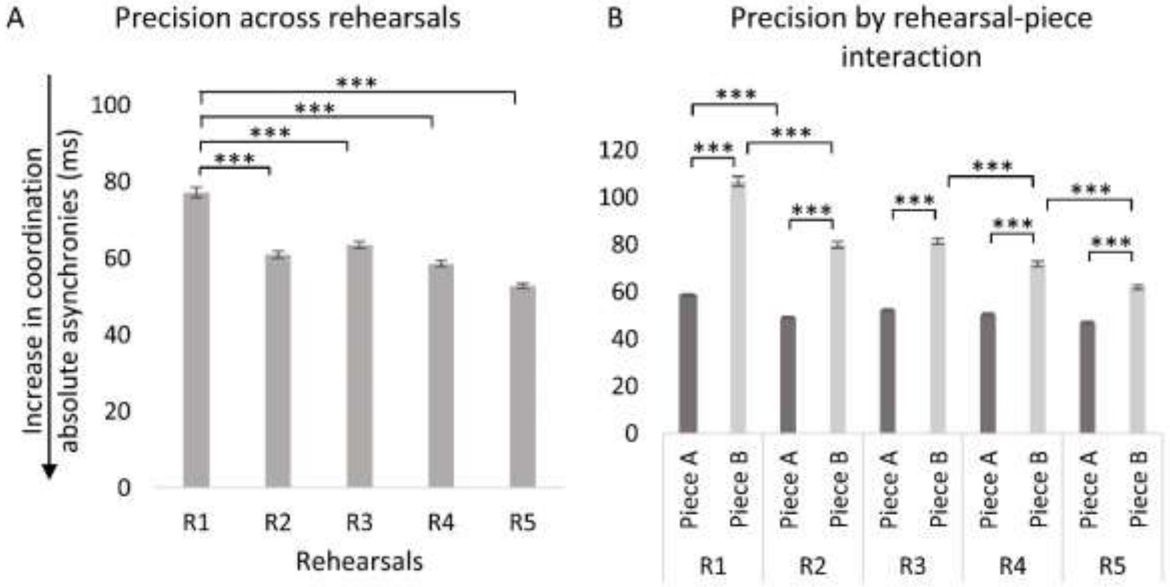


Figure 6.4 *Precision of synchronization: A) by rehearsal number, and B) by interaction between rehearsal and piece. Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $*** \ p < 0.001$.

6.3.1.2 Consistency

Consistency across rehearsals was on average 38.84 ms $(SD = 48.22\,ms)$, as computed by averaging the SD values of the absolute asynchronies collapsing the entire data set, irrespective of rehearsal number, piece, note, and time category. The five rehearsals did not predict synchronization consistency, as shown in **Figure 6.5A**, but the pieces within each rehearsal were significant predictors. The consistency in the synchronization of piece A was better than that of piece B in all rehearsals, as shown in **Table 6.4** and **Figure 5B**. The variance partition coefficient between pre- and post-rehearsal was 0.04 %, among time categories 5.6%, and among notes 12.5 %. In addition, *post-hoc* comparisons, Holm corrected for multiple comparisons, between rehearsals of the two pieces show that the consistency of piece B improved significantly from the first rehearsal ($M = 81.9\,ms, SD = 122.0\,ms$) to the second rehearsal ($M = 47.4\,ms, SD = 66.9\,ms, t = 9.3, p < 0.001$), as shown in **Figure 6.5**.

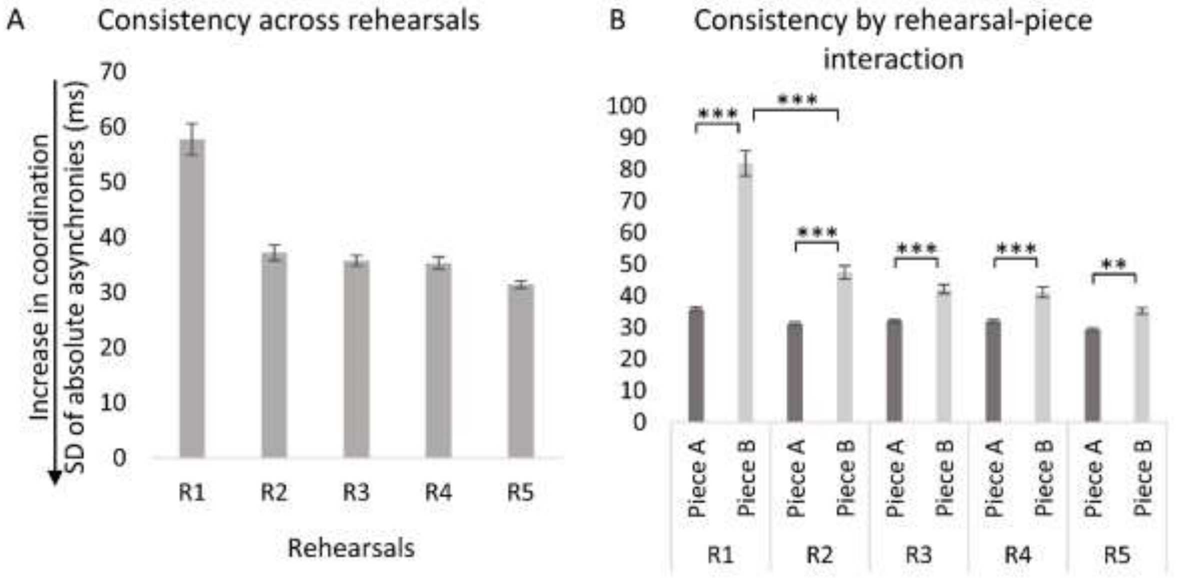


Figure 6.5 *Consistency of synchronization: A) by rehearsal number, and B) by interaction between rehearsal and piece. Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $** \, p < 0.01$; $*** \, p < 0.001$.

6.3.1.3 Tendency to Precede/Lag

The tendency to precede/lag a co-performer was 1.5 ms ($IQR = 85\,ms$), as computed by the median signed asynchronies calculated by collapsing the entire data set, irrespective of rehearsal number, piece, note, and time category. Rehearsal number did not predict the tendency to precede or lag a co-performer computed for each pair of singers as shown in **Table 6.4**. The piece being rehearsed predicted the tendency to precede/lag during the first rehearsal: the amount of leadership/lagging was greater when singers performed piece B ($M = 12.4\,ms, SD = 198.1\,ms$), than piece A [$M = 6.3\,ms, SD = 76.9\,ms$), $\beta = 10.2, t(39783) = 4.0, p < 0.001$], as shown in **Table 6.4**. The VRP among pairs was 0.9 %, notes 0.3 %, time categories 3.7 %, and between pre- and post-rehearsal 0.075 %.

In addition, Pearson's chi-square tests for each singer, which analysed the interaction between rehearsal number and positions in each rehearsal session (as defined section *6.2.6*), show that there was a significant association between the given variables. The occurrences that each singer spent in each position (P1-P5) did depend on rehearsals [for singer 1: $\chi^2(16) = 55.1, p < 0.001$; for singer 2: $\chi^2(16) = 70.2, p < 0.001;$ for singer 3: $\chi^2(16) = 63.6, p < 0.001;$ for singer 4: $\chi^2(16) = 42.8, p < 0.001;$ for singer 5: $\chi^2(16) = 54.0, p < 0.001)$]. This demonstrates that the tendency to precede/lag co-performers was significantly associated with the rehearsal sessions (i.e., R1-R5).

Figure 6.6 illustrates the time spent in each position for each singer across the rehearsals. Interestingly, the bass (S5) spent most time in position 1 in rehearsals 1-4, therefore mostly preceding all co-performers.

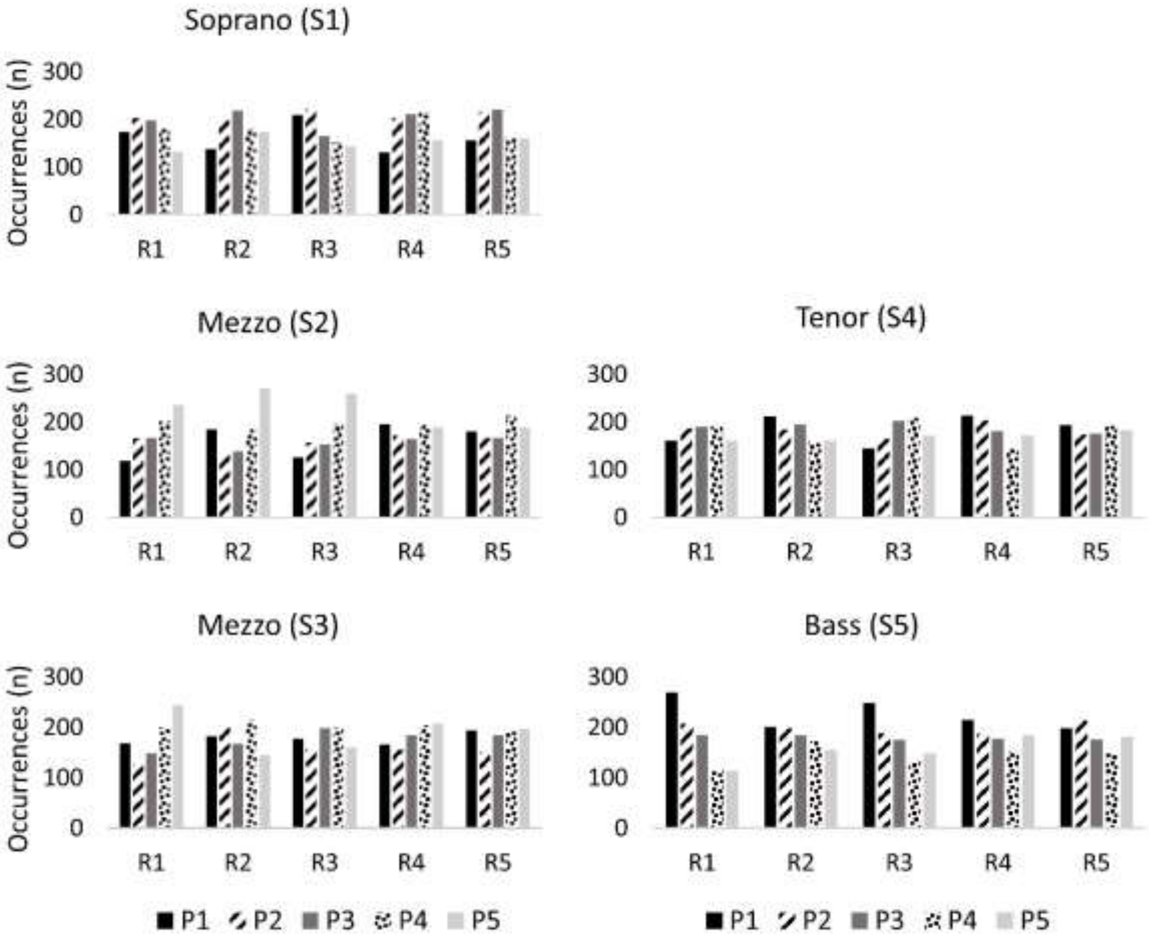


Figure 6.6 *Occurrences of entry positions from position 1 to position 5 across rehearsals computed for each singer.*

An analysis of leadership as indexed by entering in position 1 followed. This showed that the distribution between singers varied significantly across rehearsals $[\chi^2(16) = 96.7, p < 0.001]$. This result demonstrates that the tendency to precede all co-performers did relate to the different rehearsals (i.e., R1-R5). Results from the goodness of fit chi-square test indicate that the observed frequencies of position 1 for each singer (see **Figure 6.7**) were not equally distributed across rehearsal 1 $[\chi^2(4) = 69.0, p < 0.001]$, rehearsal 2 $[\chi^2(4) = 17.4, p = 0.002]$, rehearsal 3 $[\chi^2(4) = 53.1, p < 0.001]$, or rehearsal 4 $[\chi^2(4) = 27.6, p < 0.001]$. Notably, there was no significant difference between singers in occupying the first position in rehearsal 5 $[\chi^2(4) = 6.4, p = 0.172]$. This indicates that the tendency to precede all other co-performers changed during the course of study: it was not equally distributed among the five singers in the first four rehearsals, but it was during the last rehearsal, as shown in **Figure 6.7**.

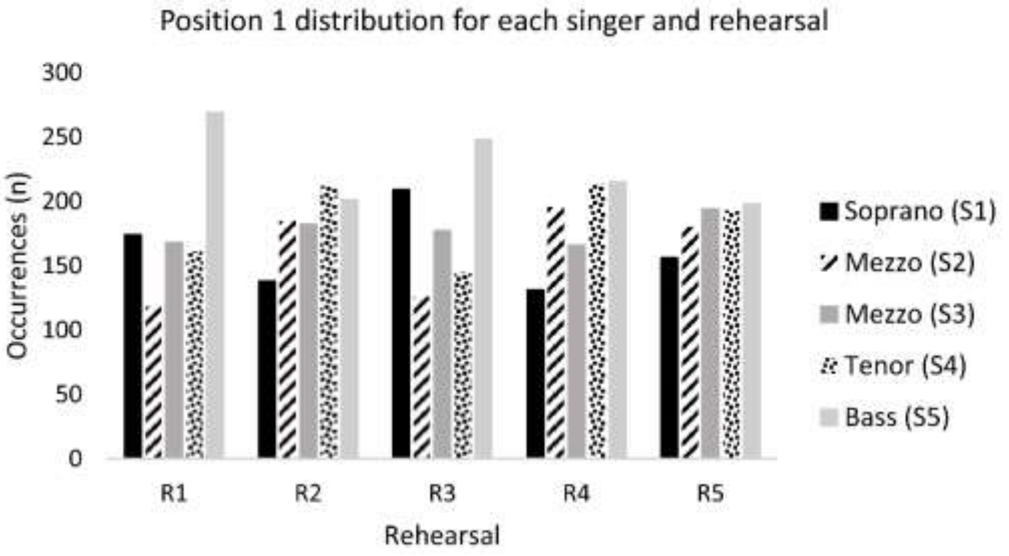


Figure 6.7 *Distribution of position 1 across rehearsals, based on the number of occurrences each singer preceded all co-performers.*

In summary, these results show that when calculated at the level of relationships between the pairs of singers, the five rehearsals were not associated with changes in these relationships. However, the tendency to precede/lag each co-performer was significantly associated with rehearsal number. Notably, the tendency to precede all co-performers became equally shared among the singers by the end of the first term of study.

Table 6.4 *Primary and nested fixed effect coefficients and significance from the analysis of interpersonal synchronization of timing. The β - fixed effect coefficients - of rehearsal number and piece on the predictor being considered (i.e., precision, consistency, and tendency to precede/lag) are given with reference to the specific base level of the factor, i.e. rehearsal 2, 3, 4 and 5 versus the base level rehearsal 1, and piece B versus piece A. The β coefficients indicate that for each 1 unit increase in the predictor being considered, the effect of the given predictor changes by the amount specified by the β coefficient. For example, for each 1 unit increase in the precision of rehearsal 1, precision computed in rehearsal 2 decreases by 9.7 units. n.s. = not statistically significant;* $*** p < 0.001$.

Parameter	**Fixed effects**	**Fixed effect coefficients and significance**				
		Rehearsal 1	*Rehearsal 2*	*Rehearsal 3*	*Rehearsal 4*	*Rehearsal 5*
Precision						
	Rehearsals		β=-9.7***, t(40505)=-5.4	β=-6.8***, t(40505)=-3.8	β=-8.8***, t(40505)=-4.9	β=-11.9***, t(40505)=-6.6
	Pieces	β=49.1***, t(40532)=23.5	β=32.4***, t(40532)=15. 5	β=30.7***, t(40532)=14.7	β=23.4***, t(40533)=11.2	β=17.7***, t(40532)=8.5
Consistency						
	Rehearsals		n.s.	n.s.	n.s.	n.s.
	Pieces	β=48.9***, t(3794)=14.1	β=19.5***, t(3798)=6.0	β=14.2***, t(3798)=4.4	β=13.4***, t(3799)=4.1	β=10.5***, t(3799)=3.2
Tendency						
	Rehearsals		n.s.	n.s.	n.s.	n.s.
	Pieces	β=10.2***, t(39783)=4.0	n.s.	n.s.	n.s.	n.s.

6.3.2 Tuning

As explained previously (see section *6.2.2*), the analysis of tuning behaviour that follows is based on piece A, but not piece B, for its clearly homophonic nature, which makes piece A of great interest to the study of tuning.

6.3.2.1 Horizontal Tuning

Visual inspection of the horizontal analysis of tuning clearly demonstrates that each singer was closer to equal temperament than just intonation, and this distinctive behaviour was consistent and repeatable before and after rehearsal, and across rehearsals (i.e., R1-R5). This is illustrated in **Figure 6.8** and **Figure 6.9**, showing the f_0 deviations computed against equal temperament and just intonation for the soprano calculated for each repeated performance in rehearsal 1 and rehearsal 5, respectively. The analysis demonstrates that the soprano tended towards equal temperament in both rehearsals and across repetitions within each rehearsal. Complete pitch-drift analysis for each singer/note/performance/rehearsal is reported in **Appendix A. Measured Tuning Deviation across Rehearsal 1**, **Appendix B. Measured Tuning Deviation across Rehearsal 2**, **Appendix C. Measured Tuning Deviation across Rehearsal 3**, **Appendix D. Measured Tuning Deviation across Rehearsal 4**, and **Appendix E. Measured Tuning Deviation across Rehearsal 5**. Based on these results, the inferential analysis of tuning during and across rehearsal was based on deviation from equal temperament, rather than just intonation.

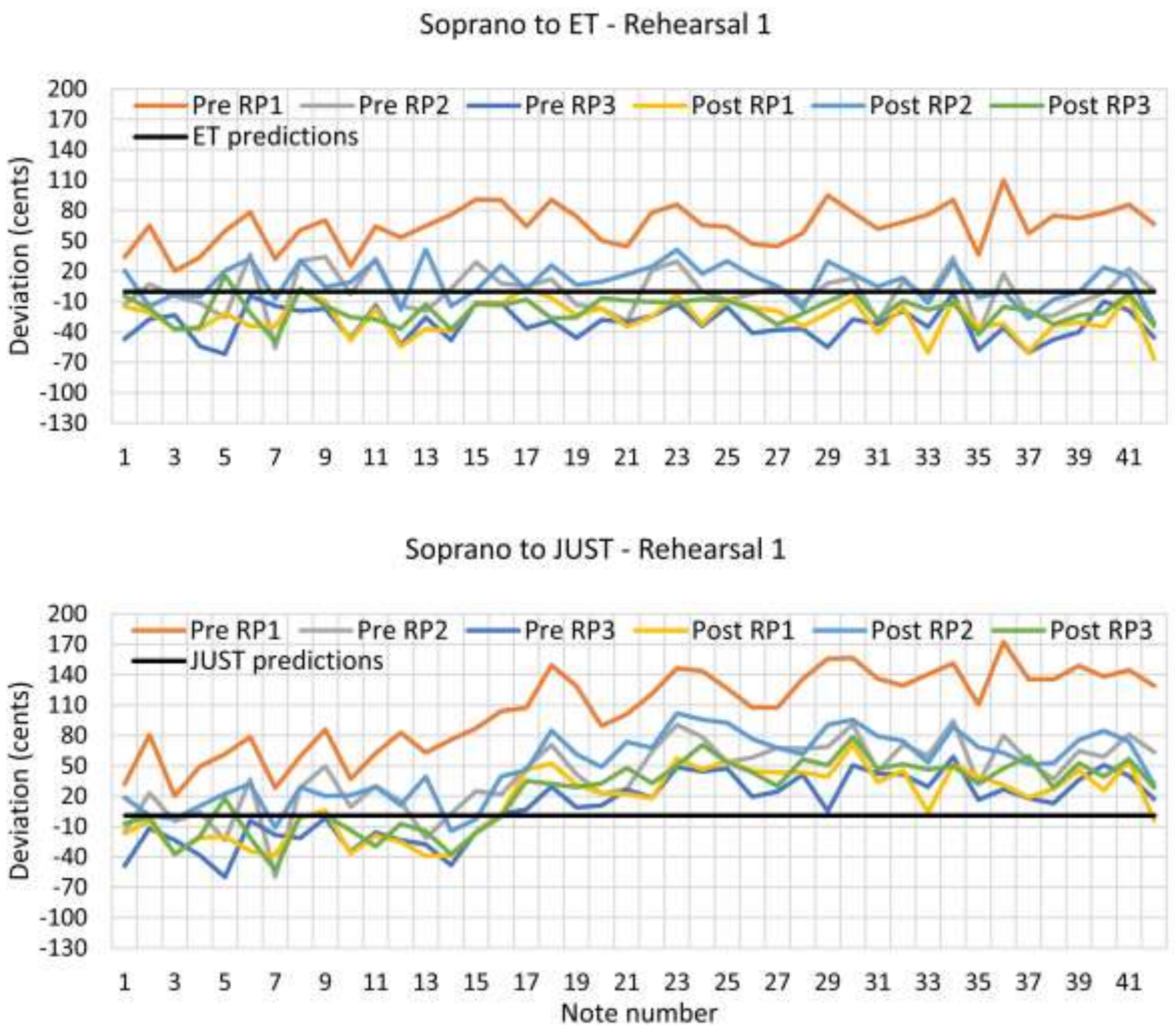


Figure 6.8 *Measured deviation from equal temperament predictions (ET, top row) and just intonation predictions (JUST, bottom) of the soprano computed for each note (notes 1-42), repeated performance (RP1-RP3), and stage (pre- and post-rehearsal) during the first rehearsal, R1. Notes are normalized to the first tenor note, F3, which is the tonic of the piece used in the study. Maximum and minimum values on the y-axis have been fixed to allow comparison between the two graphs.*

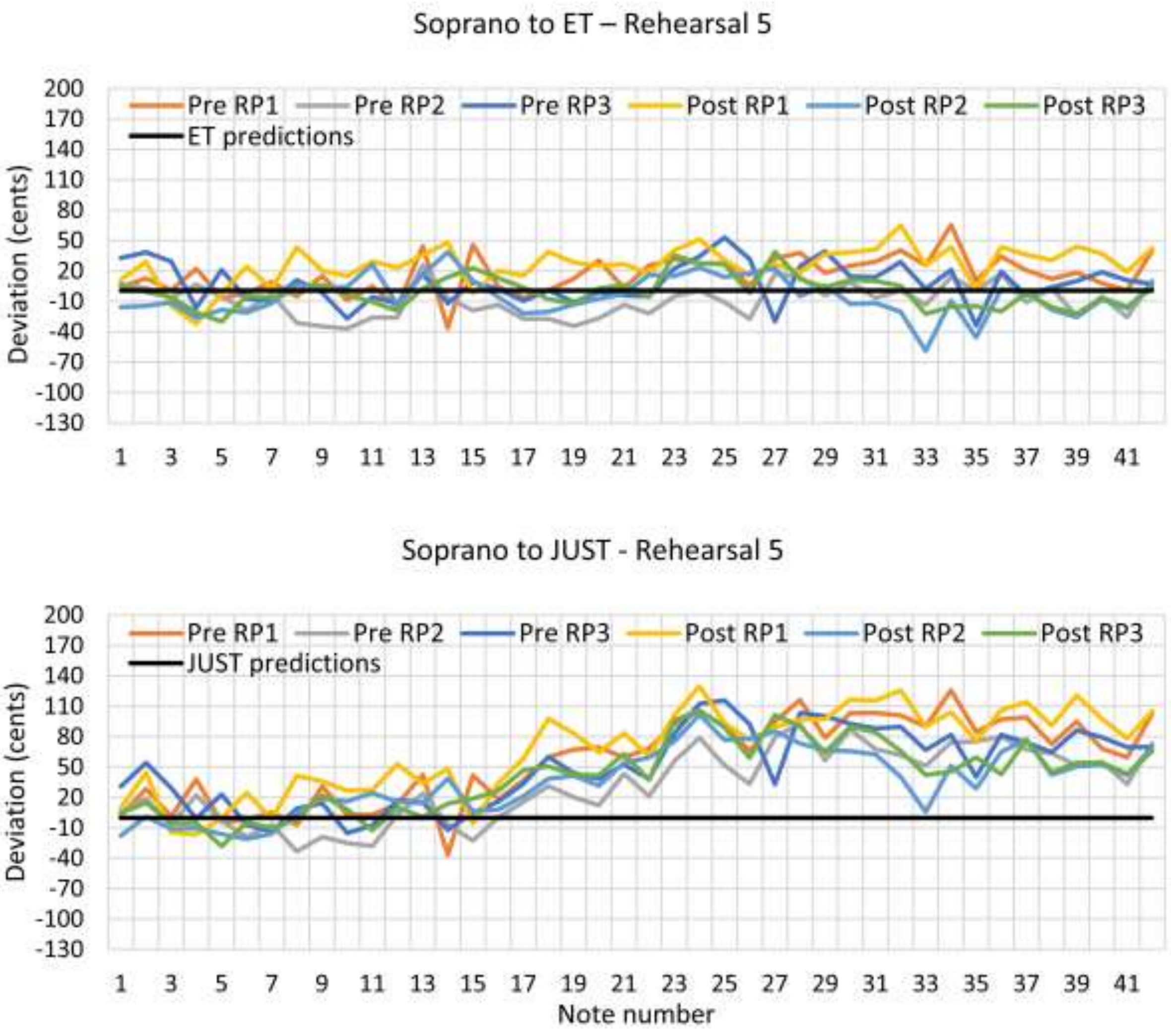


Figure 6.9 *Measured deviation from equal temperament predictions (ET, top row) and just intonation predictions (JUST, bottom) of the soprano computed for each note (notes 1-42), repeated performance (RP1-RP3), and stage (pre- and post-rehearsal) during the last rehearsal, R5. Notes are normalized to the first tenor note, F3, which is the tonic of the piece used in the study. Maximum and minimum values on the y-axis have been fixed to allow comparison between the two graphs.*

Results from the multilevel linear modelling show that, compared with rehearsal 1, the measured deviation from ET was sharper in rehearsal 2 $[\beta = 4.8, t(6120) = 3.1, p < 0.01]$, and flatter in rehearsal 4 $[\beta = -19.8, t(6120) = -12.6, p < 0.001]$ and rehearsal 5 $[\beta = -12.2, t(6120) = -7.8, p < 0.001]$, as shown in **Figure 6.10A** and **Table 6.5**. The β - fixed effect coefficients - indicate that for each 1 unit increase in the predictor being considered, the effect of the given predictor changes by the amount specified by the β coefficient. For example, for each 1 unit increase in the tuning of rehearsal 1, tuning computed in rehearsal 2 increased by 4.8 units. Deviation from equal temperament tended to be flatter post-rehearsal in rehearsals 1-4, but there was no significant difference pre- and post-rehearsal in rehearsal 5, as shown in **Figure 6.10B** and **Table 6.5**, *Post-hoc* comparisons, Holm corrected for multiple comparisons, revealed that deviation from equal temperament was flatter in rehearsal 4 $(M = -9.9\ cents, SD = 25.7\ cents)$ compared with rehearsal 3 $(M = 10.3\ cents, SD = 35.0\ cents, t = 17.9, p < 0.001)$; and, deviation from ET was sharper in rehearsal 5 $(M = 1.8\ cents, SD = 24.4\ cents)$ compared with rehearsal 4 $(M = -9.9\ cents, SD = 25.7\ cents, t = 17.9, p < 0.001)$. The variance partition coefficient (VPC) among singers and notes was 0.0206 and 0.0248, which demonstrates that only 2 % and 2.5 % of the variability of tuning can be attributed to singers and notes, respectively. The variability among repeated performances was 16.2 %, which indicates that the measured deviation from ET might have changed during repetitions. For this reason, an ANOVA test was run to test the effect of take. Results show that the take order had a significant effect $[F(2,6173) = 340.8, p < 0.001]$, and that the deviation tended to be slightly flatter across repetitions though still closer to ET, as demonstrated by the *post-hoc* comparisons using the Bonferroni correction (see **Figure 6.10C**).

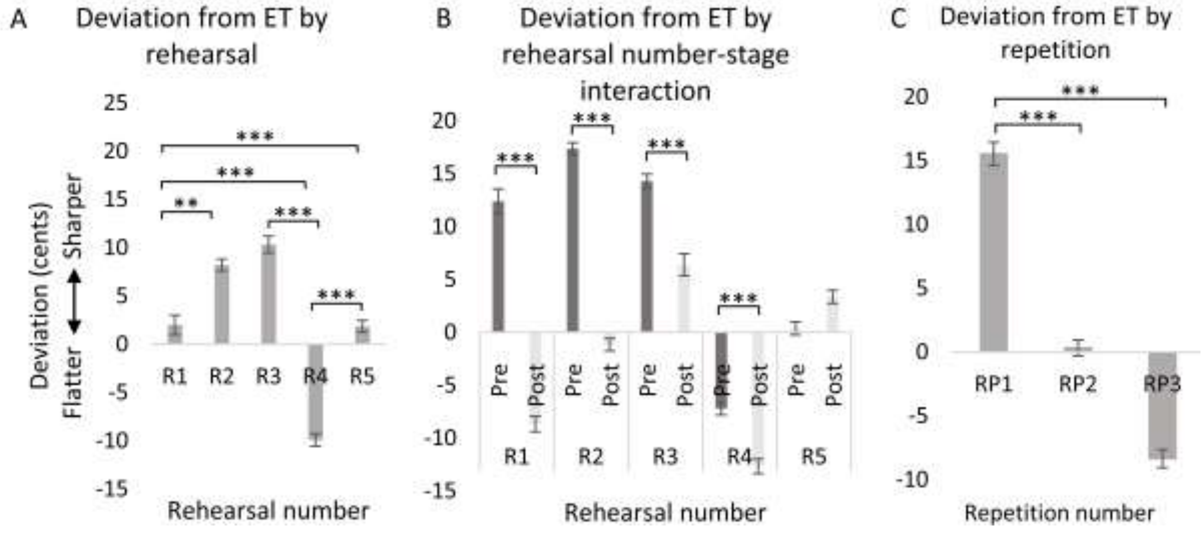


Figure 6.10 *Deviation of tuning from equal temperament predictions (ET): A) by rehearsal number (R1-R5); B) by interaction between rehearsal number (R1-R5) and rehearsal stage (pre- and post-rehearsal); and, C) by repetition from repeated performance 1 to repeated performance 3 (RP1-RP3). Error bars represent 95 % CI of the mean.* $** p < 0.01$; $*** p < 0.001$.

Results from the multilinear modelling based on the SD of measured deviation from ET show that, compared with the first rehearsal, tuning deviation was more consistent in rehearsal 2 $[\beta = -5.5, t(134) = -3.5, p < 0.001]$, rehearsal 3 $[\beta = -8.1, t(134) = -5.2, p < 0.001]$, rehearsal 4 $[\beta = -5.3, t(134) = -3.3, p < 0.01]$, and rehearsal 5 $[\beta = -4.6, t(134) = -2.9, p < 0.01]$, as shown in **Table 6.5**. Tuning was gradually more consistent during the first three rehearsals, but it did not change significantly pre- and post-rehearsal, as shown in **Figure 6.11A, Figure 6.11B,** and **Table 6.5**. The VRP among repeated performances was 6.2 %, demonstrating that the consistency of tuning did not change largely during repetitions. The VRP among repeated singers was 54.5 %, indicating that the consistency of tuning across rehearsals largely vary within the members of this quintet, suggesting that these results might change if different singers were to take part in the study. An ANOVA was run to investigate further the role of the singer in the consistency of the tuning and, as expected, results confirmed a significant effect of singer $[t(4{,}145) = 29.73, p < 0.001]$. Tuning of singer 5, the bass, $(M = 29\, cents, SD = 5\, cents)$, was less consistent compared with that of singer 1 $(M = 19\, cents, SD = 4\, cents, p < 0.001)$, singer 2 $(M = 22\, cents, SD = 6\, cents, p < 0.001)$, singer 3 $(M = 19\, cents, SD = 5\, cents, p < 0.001)$, and singer 4 $(M = 18\, cents, SD = 3\, cents, p < 0{,}001)$, as shown by Bonferroni *post-hoc* comparisons in **Figure 6.11C**.

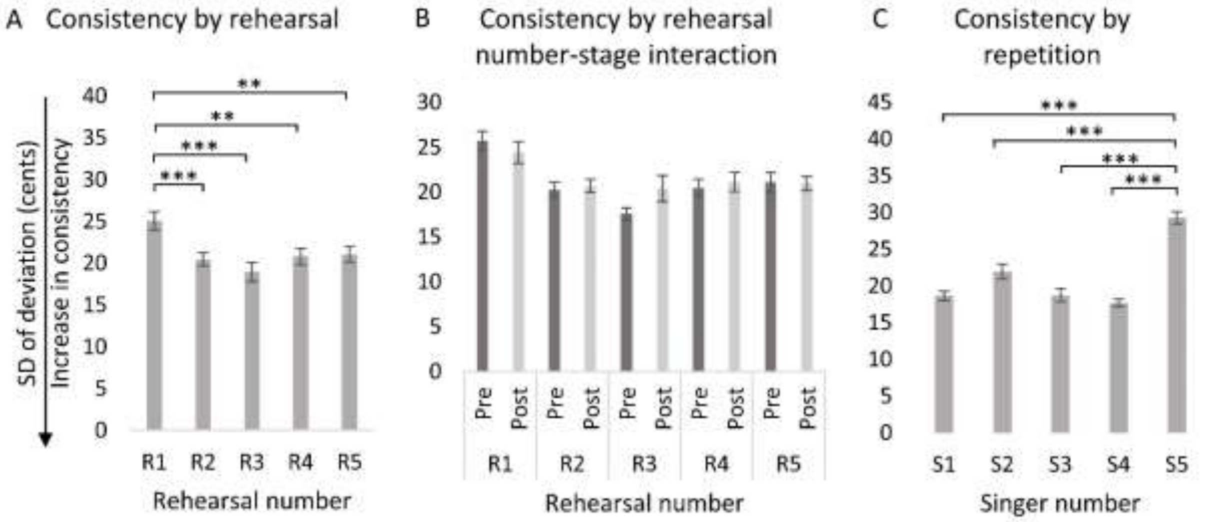


Figure 6.11 *Consistency of tuning: A) by rehearsal number (R1-R5); B) by interaction between rehearsal number (R1-R5) and rehearsal stage (pre- and post-rehearsal); and, C) by singer (S1-S5). Error bars represent 95 % CI of the mean.* $** \ p < 0.01$; $*** \ p < 0.001$.

The analysis of the dispersion of tuning across rehearsals shows that the range of tuning deviation from ET was narrower in the third rehearsal compared with the first $[\beta =$

$-33.0, t(134) = -3.7, p < 0.001]$, as shown in **Figure 6.12A** and **Table 6.5**. Tuning range did not differ significantly pre- and post-rehearsal, as shown in **Figure 6.12B** and **Table 6.5**. The variability of the primary effects of rehearsal among repeated performances and singers was 8.8 % and 38.5 % respectively, suggesting that these results might change if different singers were to take part in the study. An ANOVA was conducted to investigate further the effect of singer, and results confirmed that the dispersion differed significantly according to the singer $[t(4,145) = 16.1, p < 0.001]$. The spread of tuning was wider in singer 5, the bass, $(M = 130\ cents, SD = 28\ cents)$, compared with singer 1 $(M = 84\ cents, SD = 17\ cents, p < 0.001)$, singer 2 $(M = 99\ cents, SD = 31\ cents, p < 0.001)$, singer 3 $(M = 87\ cents, SD = 36\ cents, p < 0.001)$, and singer 4 $(M = 84\ cents, SD = 20\ cents, p < 0,001)$, as shown by Bonferroni *post-hoc* comparisons (see **Figure 6.12C**).

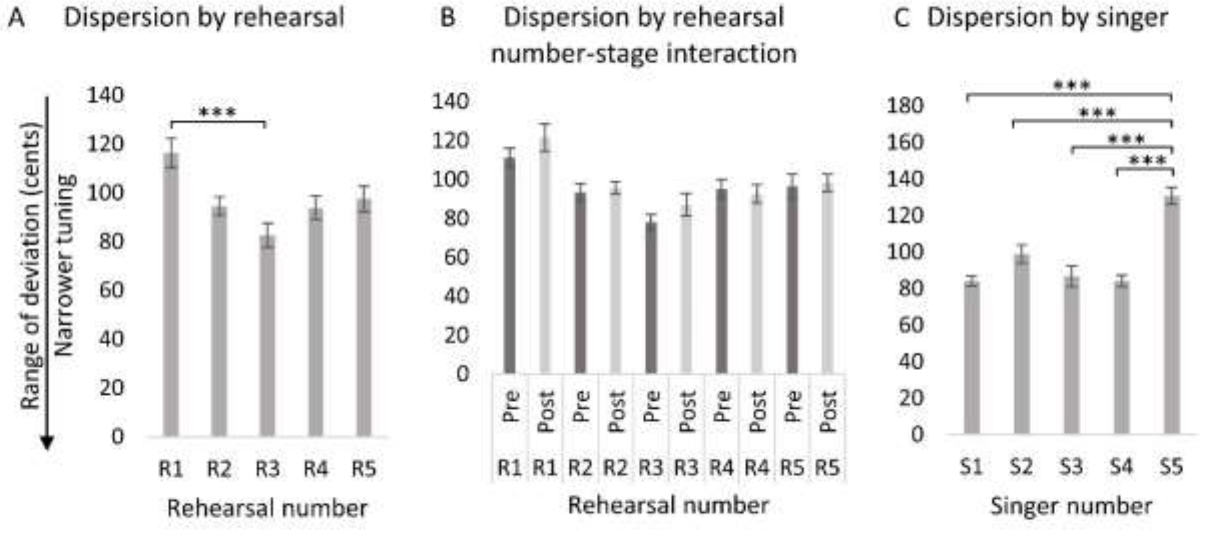


Figure 6.12 *Dispersion of tuning: A) by rehearsal number (R1-R5); B) by interaction between rehearsal number (R1-R5) and rehearsal stage (pre- and post-rehearsal); and, C) by singer (S1-S5). Error bars represent 95 % CI of the mean.* $*** p < 0.001$.

Table 6.5 *Primary and nested fixed effect coefficients and significance from the horizontal analysis of tuning. The β - fixed effect coefficients - of rehearsal number and stage on the predictor being considered (i.e., deviation from ET, consistency, range) are given with reference to the specific base level of the factor, i.e. rehearsal 2, 3, 4 and 5 versus the base level rehearsal 1, and post- versus pre-rehearsal. The β coefficients indicate that for each 1 unit increase in the predictor being considered, the effect of the given predictor changes by the amount specified by the β coefficient. For example, for each 1 unit increase in the consistency of rehearsal 1, precision computed in rehearsal 2 decreases by 5.5 units.* $n.s$ *=not statistically significant;* $\ast\ast\ p < 0.01$; $\ast\ast\ast\ p < 0.001$.

Parameter - **Fixed effects**	**Fixed effect coefficients and significance**				
	Rehearsal 1	*Rehearsal 2*	*Rehearsal 3*	*Rehearsal 4*	*Rehearsal 5*
Deviation					
- Rehearsals		$\beta = 4.8\ \ast\ast$, $t(6120) = 3.1$	n.s.	$\beta = -19.8\ \ast\ast\ast$, $t(6120) = -12.6$	$\beta = -12.2\ \ast\ast\ast$, $t(6120) = -7.8$
- Stage	$\beta = -21.3\ \ast\ast\ast$, $t(6120) = -13.4$	$\beta = -18.5\ \ast\ast\ast$, $t(6119) = -11.9$	$\beta = -7.9\ \ast\ast\ast$, $t(6119) = -5.0$	$\beta = -5.5\ \ast\ast\ast$, $t(6119) = -3.5$	n.s.
Consistency					
- Rehearsals		$\beta = -5.5\ \ast\ast\ast$, $t(134) = -3.5$	$\beta = -8.1\ \ast\ast\ast$, $t(134) = -5.2$	$\beta = -5.3\ \ast\ast$, $t(134) = -3.3$	$\beta = -4.6\ \ast\ast$, $t(134) = -2.9$
- Stage	n.s.	n.s.	n.s.	n.s.	n.s.
Range					
- Rehearsals		n.s.	$\beta = -33.0\ \ast\ast\ast$, $t(134) = -3.7$	n.s.	n.s.
- Stage	n.s.	n.s.	n.s.	n.s.	n.s.

6.3.2.2 Vertical Tuning

The average size of the major thirds was 392.17 cents with a standard deviation of 27.56 cents. This is slightly closer to just intonation (386 cents) than ET (400 cents), and, together with the wide spread, indicates examples of both ET and just intonation, as shown in **Figure 6.13A**. The stability of the thirds did not change significantly across rehearsals or pre- and post-rehearsal. The pairs of singers involved in the performance of major thirds were S1-S2, S3-S2, S4-S3, and S5-S4. The variability among interval number, pair and repeated performances was 5.2 %, 5.9 % and less than 0.1 %, respectively. Considering the significant effect of singer on the horizontal tuning, an ANOVA was conducted to test whether tuning of the thirds changed according to the pair of singers performing. Results demonstrate a significant effect of the pair of singers $[t(3{,}438) = 9.0, p < 0.001]$, which was associated with the pair S3-S2 and S1-S2, as shown by Bonferroni *post-hoc* comparisons (see **Figure 6.13B**). The pair S1-S2 tuned the major thirds closer to just intonation, but the pair S2-S3 tuned closer to ET. Another ANOVA was also conducted to test the effect of note number, and results show the major thirds tuning changed significantly based on the chord considered $[t(14{,}427) = 4.1, p < 0.001]$.

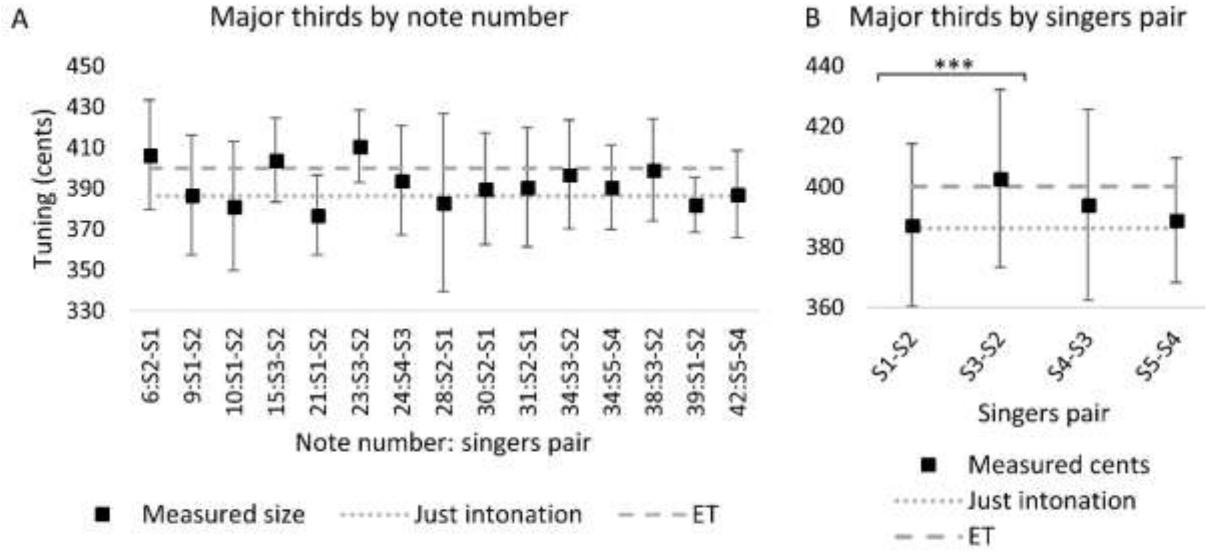


Figure 6.13 *Tuning of major thirds: A) by note number, and B) by singers' pair. Error bars represent SD of the mean.* $*** \ p < 0.001$.

The consistency and range of the major thirds did not differ across rehearsals or pre- and post-rehearsal, and the variance partition coefficient among repetitions was 8 % in relation to the consistency and 15.2 % for the range of major thirds. Further tests were then conducted to investigate the role of the repeated performances, and ANOVAs show that the consistency and dispersion of tuning of the major thirds did not differ across repetitions.

The average size of the minor thirds was 299.13 cents, with a standard deviation of 29.28 cents, indicating that the tuning of the minor thirds was closer to ET (300 cents), than just intonation (315.6 cents), as shown in **Figure 6.14**. The tuning stability, as indexed by the size of interval, did not differ pre- and post-rehearsal, or across rehearsals. The variability among minor thirds, pair and repetitions was 15.4 %, 1.3 % and less than 0.1 %, respectively. An ANOVA on the minor thirds number confirmed a significant effect of the interval number on the tuning of the minor thirds $[t(22{,}646) = 6.4, p < 0.001]$. The consistency and range of the minor thirds did not change across the five-rehearsal sessions or pre- and post-rehearsal, and the variance partition coefficient among repeated performances 1-3 was less than 0.1 % in relation to both consistency and range.

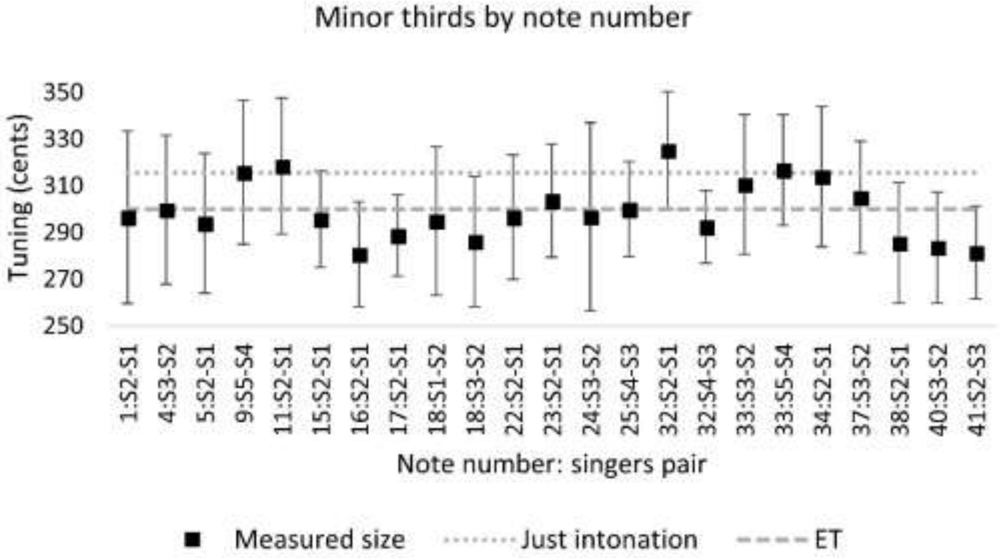


Figure 6.14 *Tuning of minor thirds by note number.*

6.3.3 Verbal Interactions

The scrutiny of the verbal discussions in relation to specific reference to synchronization shows that singers never discussed nor debated synchronization during the five rehearsal sessions. On the other hand, there were discussions related to tuning. Over the five rehearsal sessions singers allocated 19 % of their rehearsal time to talking about tuning, and the allocation time to tuning discussions decreased across rehearsals, as shown in **Figure 6.15**.

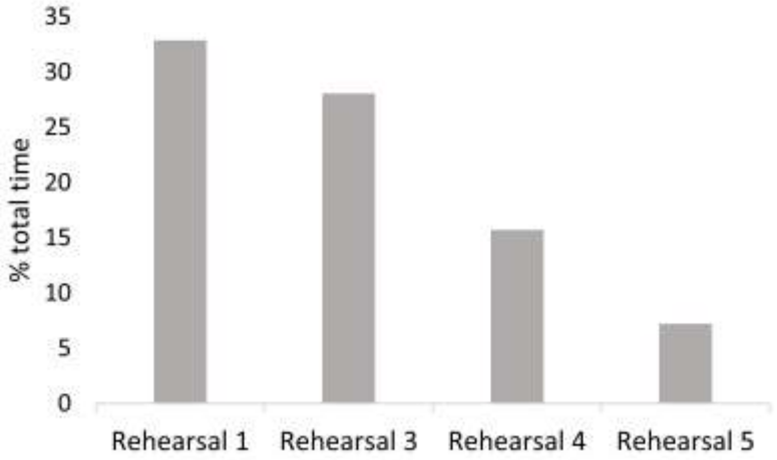


Figure 6.15 *Allocation time for tuning discussions of each rehearsal expressed as percentage of the total rehearsal time of the related rehearsal.*

The amount of time spent talking about tuning is summarised in **Figure 6.16** by bar number.

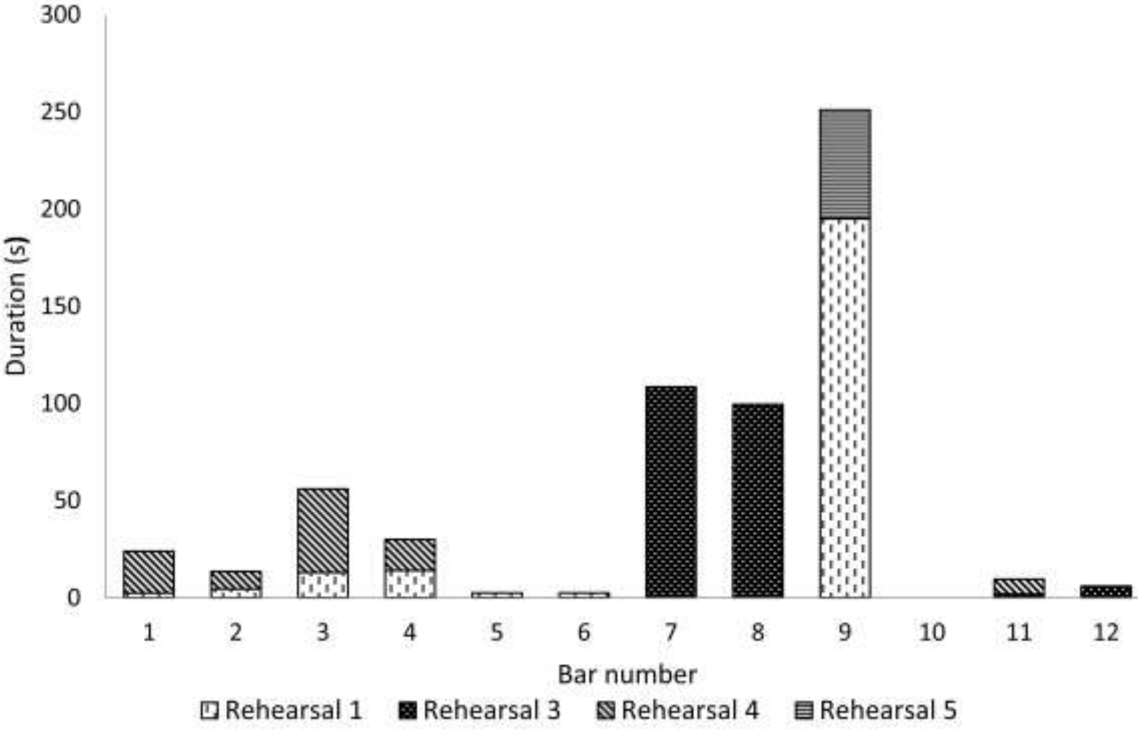


Figure 6.16 *Time spent talking about tuning by bar number for each rehearsal.*

Figure 6.17 displays the time spent on tuning by chord number. At least 20 s were spent on chords n 30, 32, 10, 24, 26 for tuning related reasons, suggesting that these might have been more challenging to tune than the others.

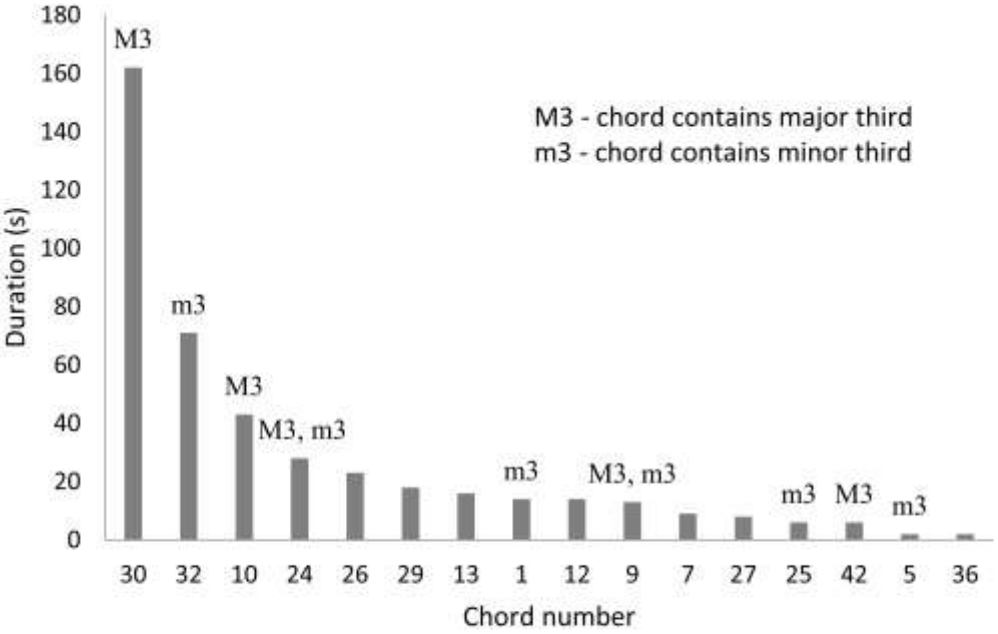


Figure 6.17 *Time dedicated to tuning, showing the chords of greatest interest from left to right. Chords of interest containing major and/or minor thirds are also highlighted. Chords to which the ensemble did not allocate any time to work on tuning, are not reported.*

Most of the time spent on tuning was related to chords involving intervals of major thirds (i.e., chords n 30, 10, 24, 9, and 42); less time was spent tuning chords involving minor thirds (i.e., chords n 32, 24, 1, 9, 25, 5), as shown in **Figure 6.18**. Only a small amount of time was spent on tuning chords that do not include either major or minor thirds intervals (i.e., chords n 26, 29, 13, 12, 7, 27, 36), although the total number of occurrences of these chords (14) is similar to the total number of major thirds (15) occurring in this piece.

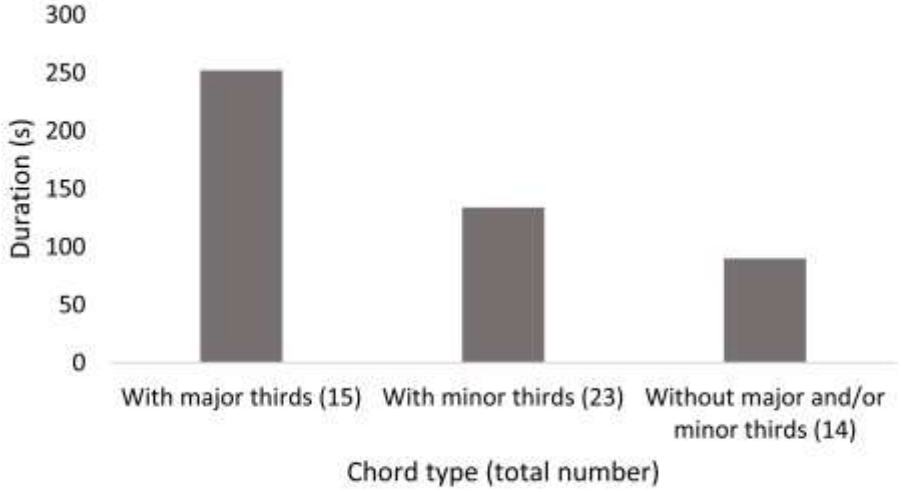


Figure 6.18 *Amount of time spent tuning chords including or not major and/or minor chords; in () the total number of chords per type occurring in the piece (i.e., Piece A).*

Based on the analysis of the verbal interactions, only one case was an explicit reference made (by singer 4) to the tuning system adopted; he stated:

> *I'm going to do an equal tempered third at the end this time (Singer 4, chord 42, rehearsal 5)*

But there was no other discussion or direct reference to tuning systems during rehearsal.

During rehearsals, the attention was focused on how to tune double notes in the same chord, and "fuzzy" chords; and, the following strategies were employed to solve tuning-related issues:

- Running or repeating a short section, single bar or chord.
- Separating out parts so that just two or three voices could be isolated.
- Singing a progression more slowly, encouraging each other to listen in a more focused way.
- Rebalancing chords so that certain voices could be stronger.

In very few cases the singers made an explicit reference to how they plan to adjust their tuning. For example, singer 3 in two cases declared she was going to lower a minor third, and also invited singer 2 to do the same.

> *I was trying to pull the F down. (Singer 3, chords 23, 25 and 27, rehearsal 3)*

> *I think it might settle if you really make that semitone close, so you can sit down lower. (Singer 3 to Singer 2, chord 30, rehearsal 1)*

As far as chords containing major and minor thirds, singers made use of a number of tuning strategies, such as balancing voices, and tuning "double" notes, as shown in **Table 6.6**. There were some examples where singers referred to the tuning of specific melodic or harmonic thirds. Singer 2 expresses difficulty with the tuning of melodic major thirds in bars 1 and 11:

> *I'm very conscious of my falling major thirds in bar 1 and bar 11, I'm finding them quite hard to tune, I don't know why. (Singer 2, rehearsal 4, chords 1 to 4 and 36 to 39).*

In rehearsal 3, singer 4 draws attention to a harmonic minor third, which prompts singers 3 and 5 to work on tuning of chord 24:

> *I think I'm hearing the minor third between the F and the D, [singer 3 and singer 5] as too wide...in other words the D on the third beat of the bar as too low. (Singer 4, chord 24, rehearsal 3)*

Table 6.6 *Tuning strategies for major and minor thirds.*

Tuning strategies	**Major thirds chord number (voices)**	**Minor thirds chord number (voices)**
Tuning ‘doubled’ notes	9 (S1, S2) 10 (S1, S2) 24 (S3, S4) 42 (S4, S3/S2)	9 (S5, S4) 24 (S3, S2) 25 (S4, S3) 32 (S4, S3)
Balancing voices	30 (S1, S2)	5 (S2, S1) 32 (S2, S1)
Tuning of whole chord	9 (S1, S2) 10 (S1, S2) 30 (S1, S2)	1 (S2, S1) 32 (S4, S3)
Tuning melodic intervals		1 (to 2) (S2) 24 and 25 (S3, S5)
Aiming for equal temperament	42 (S4, S5)	

In summary, the results from the analysis of the verbal interactions show that synchronization was not a topic of discussion, whilst tuning was an important dimension of rehearsal, whose focus decreased across rehearsals. Singers made use of a range of strategies to solve tuning related issues, including tuning doubled notes, whole chords, specific melodic intervals, and balancing voices, but, except for one instance, there was no explicit reference to the tuning system adopted.

6.3.4 Overview: Synchronization-Tuning-Verbal Interactions

A summary of the analyses of synchronization, tuning, and verbal interaction is presented in **Table 6.7** to provide a visual snapshot of the entirety of the longitudinal pairings between the multiple parameters and aspects analysed in the previous three sections. Notably, as the table highlights, most of the significant changes happened in the second rehearsal compared with the first. Specifically, compared with rehearsal 1, rehearsal 2 features:

- An improvement
    - In the precision of synchronization of both pieces
    - In the consistency of synchronization and tuning of piece B
- A tendency to tune sharper horizontally in piece A

Other changes were observed also in the last rehearsal, the fifth, compared with the fourth. During the last two rehearsals, the precision of synchronization of piece A improved and in parallel results show a shift from a non-equal to equal distribution of the tendency to precede all co-performers, paired with a varying distribution of the verbal utterances between singers.

Table 6.7 *Overview of the significant changes regarding synchronization and tuning by rehearsal computed by comparing a given parameter with that of the precedent rehearsal (e.g., precision of rehearsal 2 compared with that of rehearsal 1.) alongside the analysis of the verbal discussions and tendency to precede/lag per rehearsal.* ↑ *=improvement in the synchronization or tuning;* ↑ *=tuning sharper than ET predictions;* ↓ *= tuning flatter than ET predictions; n.a. =not available.*

Data set	**Parameter**	**Piece**	**R 1**	**R 2**	**R 3**	**R 4**	**R 5**
Synchronization	Precision	A		↑			
	Precision	B		↑		↑	↑
	Consistency	A					
	Consistency	B		↑			
	Tendency to precede/lag	A and B	Not equally shared	Not equally shared	Not equally shared	Not equally shared	Equally shared
Verbal discussions	Synchronization (%)	A and B	n.a.	n.a.	n.a.	n.a.	n.a.
	Tuning (%)	A	33	n.a.	28	16	7
Tuning	Horizontal deviation	A		↑		↓	↑
	Horizontal consistency	A		↑			
	Horizontal Range	A					
	Major 3rds stability	A					
	Major 3rds consistency	A					
	Major 3rds Range	A					
	Minor 3rds stability	A					
	Minor 3rds consistency	A					
	Minor 3rds Range	A					

6.4 Discussion

This investigation analysed the evolution of interpersonal synchronization and leader-follower relationships in a newly formed *a cappella* singing quintet across five rehearsal sessions in relation to the complexity of the pieces performed. In addition, the study of synchronization and leadership was conducted in parallel with aspects of tuning and verbal interactions, to contextualise the results of synchronization within the wider scope of expressive performance goals.

6.4.1 Synchronization

The developmental aspects of timing synchronization were investigated during and across rehearsals, and in relation to the pieces practised. Three measures of interpersonal synchronization were investigated: precision and consistency of synchronization, as quantified by the absolute and SD of absolute asynchronies, and tendency to precede or lag a co-performer, as indicated by the signed asynchronies. These measures were objectively quantified through the analysis of the acoustic and Lx recordings.

Precision significantly improved from the first to the last rehearsal. In each rehearsal, precision was better in piece A, the more homophonic piece, than piece B (more polyphonic). Notably, precision in piece A improved significantly only between the first two rehearsals, but improved across the whole term of study in piece B. This suggests that the complexity of the piece being practised might affect the precision of synchronization between performers in ensembles. Singers practising a homophonic piece might significantly improve the precision of interpersonal coordination with only two rehearsals, establishing a stable degree of synchronization for the remaining rehearsals. Conversely, with a more complex piece, performers might need several rehearsals to establish a stable degree of synchronization. Further studies that increase the number of rehearsals analysed will inform whether/when singers establish a higher degree of synchronization in piece B, the more complex material.

Synchronization in piece A was more consistent than piece B in each rehearsal, as quantified by the SD of absolute asynchronies. The consistency of synchronization did not change in piece A during the full term of study. The consistency improved significantly between the first two rehearsals of piece B and then remained stable during the remaining rehearsals.

Analyses show that while singers varied in the balance of leadership (as indexed by preceding all other voices), across the first four rehearsals, by the final rehearsal no significant differences between the members of the quintet were apparent in occupying the first position. Although the analysis of the rank order positions does not offer a thorough analysis of leader-follower relationships, these results show that the tendency to precede all co-performers changes across rehearsals, becoming equally distributed among singers toward the end of a first-term of study. These results further expand findings based on single laboratory sessions, suggesting a complex pattern of relationships between string players in ensemble quartets, rather than a clearer separation of roles (Timmers et al. 2014; Timmers, Endo, and Wing 2013). The previous investigations provided a single snapshot of leader-follower relationships in string quartets, reporting: i) a unidirectional dependence of Viola on Violin I, and of Violin I on Cello; and, ii) a bidirectional dependence between Violin II and Cello, and Violin II and Viola. This study sheds some light on the developmental aspects of the group relationships in music ensembles, showing a shift from a non-equal distributed tendency to precede all other co-performers in early rehearsals, to a tendency to precede all co-performers equally shared among the five singers by the end of the first term of study. In addition, these results also expand findings based on qualitative interviews among members of a professional singing octet and an 11-men singing ensemble, reporting a shared leadership (conceptualized as social roles) among singers (Lim 2014; Page-Shipp, Joseph, and van Niekerk 2018).

In addition, the above findings regarding the relationship between rehearsal number and synchronization did not vary largely among pairs of singers and time categories. Notably, synchronization results were also consistent between pre- and post-rehearsal, suggesting that an individual rehearsal might not affect the synchronization. Singers were not told to focus on synchronization, but on expressiveness; they may improve precision and consistency of synchronization in different rehearsals if temporal coordination is the goal of the rehearsal.

Finally, the main synchronization characteristics reported across rehearsals regardless the piece practised (i.e., mean absolute asynchronies of 62.69 ms; average SD asynchronies of 38.84 ms; and, median signed asynchronies of 1.5 ms) are in line with the literature investigating synchronization during instrumental ensemble performances, reporting typical mean absolute and SD asynchronies falling between 30 and 50 ms, and mean signed asynchronies often close to 0 ms (Keller 2014; Goebl and Palmer 2009; Timmers et al. 2014; Bishop and Goebl 2015; Zamm, Pfordresher, and Palmer 2015).

6.4.2 Tuning

In addition to interpersonal synchronization, another fundamental aspect of an expressive performance was also investigated in this chapter, tuning behaviours, in order to contextualise the results from the synchronization data within the wide scope of musical expressive possibilities. Tuning was investigated by analysing the f_0 deviation measured against equal temperament and just intonation predictions, and calculated both horizontally (i.e. for each note/repetition/singer/rehearsal) and vertically (i.e., in relation to the major and minor thirds of the piece).

Each singer in this study consistently tended towards equal temperament during and across rehearsals. These results corroborate previous investigations conducted by Devaney, Mandel, and Fujinaga (2012), showing no evidence of pitch drift in a three-part exercise composed by Benedetti. However, they contrast with findings from Howard (2007a; 2007b), who observed pitch drift in a four-part piece, composed for the study, in which each chord was linked via a tied note. These findings suggest that tuning in singing ensembles might depend on the specific melodic/harmonic characteristics of the piece being performed as well as the individual singers and combination of singers performing.

Furthermore, compared with the first rehearsal, intonation computed against ET was significantly flatter in rehearsal 4 and 5, i.e. towards the end of the first term of study. It was also flatter with repeated performances, and, in most rehearsals, post-rehearsal. Tuning deviation from ET was less consistent in the first rehearsal compared with the remaining rehearsals, as shown by the SD of the measured deviation. This is not surprising, as the singers did not know each other before the first rehearsal, and had never seen the piece before. The consistency did not change pre- and post-rehearsal, but gradually improved in the first three rehearsals. The tuning of singer 5, the bass, was significantly less consistent and wider compared with the other singers, as quantified by the SD and range of measured deviation, respectively, but still closer to equal temperament than just intonation. These peculiar characteristics observed in the tuning for singer 5 could be due to the physiology of the bass voice, whose vocal folds vibrate slower than the other voices singing at higher frequencies (Johan Sundberg, personal communication, June 2018).

Tuning deviation from ET was more consistent and narrower in the third rehearsal, which was the anticipated midpoint of the first term of study, although, due to some last-minute issues, the final rehearsal date was moved, and consequently the anticipated and actual midpoint were different. Therefore, rehearsal 3 was at the time of the anticipated midpoint of the rehearsal schedule, and featured distinctive tuning outcomes in terms of the consistency and dispersion of intonation, implying a peculiar role of rehearsal 3. These results can be understood in light of the group development theory advanced by Gersick (1988), suggesting a turning point in the development, halfway through the process of working towards a shared goal, in which there is a transition from exploration mode to action planning mode.

The size of the major thirds was slightly closer to just intonation, with examples of both just and ET system across the piece. This did not change within (i.e., pre- and post-rehearsal) and across rehearsals (R1-R5), and repetitions (RP1-RP3). Chord number and pair, however, did significantly affect the size of the major thirds. The pair singer 1 and singer 2 (S1-S2) tuned their major thirds closer to just intonation, but the pair singer 2 and 3 (S2-S3) closer to ET. These results suggest that the tuning of major thirds might change with singer and the harmonic/melodic content of the piece. The highly variable size of the major thirds across pairs of singers measured in this study also corroborates the results from perception studies, showing different subjective preferences for the size of major third dyads (Ternström and Nordmark 1996).

Intonation of the minor thirds was clearly closer to ET; this did not change across repetitions and rehearsals, or pre - post rehearsal. Chord number significantly affected the tuning, suggesting, similarly to the finding with the major thirds, that intonation of the minor thirds might be context-specific. The variability of the minor and major thirds based on the chord number, including examples of intonation close to both ET and just intonation, is in line with previous results found by Devaney, Mandel, and Fujinaga (2012) when investigating minor and major thirds in a three-part progression written by Benedetti. The consistency and distribution of the major and minor thirds' tuning, as quantified by the SD and range of the thirds' size respectively, did not differ pre- and post-rehearsal, and across rehearsals and repetitions, suggesting that this tuning behaviour was highly repeatable in relation to the minor and major thirds and typified the ensemble.

6.4.3 Expressive Ensemble Performance: Combining Synchronization – Tuning – Verbal Interaction Data

The combination of measured synchronization data and verbal interactions demonstrates that the precision and consistency of synchronization changed over the study period, depending on the piece performed, but these changes were not the result of verbal debate, as no rehearsal strategies targeting note synchronization were measured. This was demonstrated, in fact, by the lack of specific reference to synchronization in the verbal discussions during the five rehearsals.

Notably, the triangulation of measured synchronization, tuning, and verbal interaction data from piece A, the clear homophonic piece, reveals a complex picture of the development of expressive performance. The improvements in the precision and consistency of synchronization observed between the first two rehearsals, although not linked to explicitly mentioned synchronization-related strategies, are paired with a complex approach to tuning, supported by a number of discussed rehearsal strategies employed to solve tuning-related issues. These synchronization patterns across rehearsals are paired with an overall horizontal tendency to ET, with a flattening towards the end of the study period, and features of just intonation and ET occurring frequently in the vertical tuning of major and minor thirds, respectively.

Notably, most of the changes were observed between the first two rehearsals during which precision of synchronization and consistency in the horizontal tendency to ET improved in piece A, alongside an improvement in the precision and consistency of synchronization in piece B. These results can be understood in light of the fact that the singers did not know each other before the first rehearsal, and had never practised the piece before. Other changes were observed during the last two rehearsals, featuring an improving in the precision of synchronization in piece B paired with a shift from a non-equally to equally distributed tendency to precede all co-performers. These results might be related to the fact that the last rehearsal took place two days before the ensemble members' formal performance exam, set up in the form of a public concert, so when the ensemble should have performed at its best.

The analysis of the verbal interactions also serves to contextualise other aspects of the tuning outcomes. The distinctive horizontal tendency to tune closer to ET (which was consistent and repeatable before and after rehearsal and across rehearsals) in parallel with the increased consistency in the horizontal tuning during the first three rehearsals and the lack of longitudinal changes in the tuning of the thirds is unexpected. However, these results suggest that the group is satisfied with these intonation outcomes, also implied by the reduced discussion on tuning in the later sessions. The varied tuning of thirds, including examples of both ET and just intonation, consistently across repetitions, rehearsal number, and stage of rehearsals (i.e, pre - post rehearsal), suggests either context and/or singer specific practice. However, this seems to be a spontaneously emerging characteristic based on the absence of specific work to tune thirds, rather than the group worked to tune an overall chord or match octaves. The notable, larger amount of time spent on the tuning of chords with major thirds compared with the remaining chords observed in the current study is in line with experience reported in the literature that thirds are difficult to tune (Covey-Crump 1992). However, without explicit reference to the thirds within the verbal discussions between singers in this quintet, these results support the hypothesis that tuning is highly context-driven, based on a complex combination of factors, rather than a simple aim to tune thirds within a specific tuning system.

The adherence to a specific tuning system was not debated during the rehearsal sessions, except for only one instance in which singer 4 declared the intention to tune the last chord of piece A to ET. Further interviews should be conducted to investigate the singers' intentions for aspects related to tuning and synchronization, to shed some light on the conscious decisions of this ensemble, which might not have emerged in the verbal discussions.

Differences in the synchronization and tuning across the advanced singing students' first term of study as an ensemble might be related to a number of different extraneous variables. These changes are most probably related to other performance goals including blend and expression, and reflect the findings of the verbal interaction data that they spend less time discussing tuning throughout the term. External factors occurring between rehearsals might have elicited an improvement in the synchronization and in the consistency of the overall horizontal tuning. Singers were rehearsing the Masters degree recital pieces between the five rehearsal sessions that took place in laboratory, and the more time spent together as an established ensemble during the course of the study might have elicited an improvement in synchronization and tuning. Further studies should analyse the role of rehearsal in more depth to understand whether the rehearsal has a significant effect on the synchronization between musicians and in the tuning behaviours manifested in this ensemble.

6.5 Limitations and Future Work

This study focused on five laboratory-based rehearsal sessions, representing five snapshots captured across a first term of study. The ensemble continued to rehearse outside of the study in order to work on other pieces, and was coached between laboratory-based rehearsals; these extra events were not considered in this study but will have contributed to the development of the group's performance traits. Further research should analyse all rehearsal and coached sessions of a specific and shorter study period to investigate the development of tuning with controlled practice, and in relation to coaching.

The intention of the current investigation was to analyse aspects of synchronization and tuning that emerged spontaneously during rehearsal. For this reason, singers were not aware of the focus of this study, but were asked to work on producing an expressive performance. Another avenue of research should consider the evolution of synchronization and tuning when singers are explicitly asked to master such aspects. This may shed some more light on the rehearsal strategies that singers consciously apply to develop tuning and synchronization during singing ensemble performances, as well as determine which tuning systems they are aspiring to.

In addition, singers were invited to master the expressive performance of the pieces across rehearsals, pretending that they would have had to perform the piece in the form of a concert. Further studies should consider a more realistic situation, with an ensemble working on a piece that will be then also performed on stage.

6.6 Conclusion

This chapter has described a longitudinal study of synchronization between singers in a semi-professional singing quintet across five rehearsal sessions during a four-month study period, to investigate the developmental aspects of synchronization and leader-follower relationships, in relation to the rhythmical complexity of the pieces practised. In addition, the study of tuning behaviours and verbal interactions during rehearsals has been conducted and results discussed, to contextualise synchronization within the broader scope of expressive performance behaviours.

Precision of synchronization increased across rehearsals depending on the piece being rehearsed: it improved between the first two rehearsals in the less complex piece, piece A, and consistently across the five rehearsals in the most complex piece, piece B. Likewise, consistency of synchronization was piece-specific: it did not change during the first term of study in piece A, but improved in the first two rehearsals in piece B. These synchronization patterns observed were not associated with any acknowledged rehearsal strategy adopted to improve synchronization.

The tendency for members of the ensemble to precede/lag the others became equally shared among the five singers by the last rehearsal session, suggesting a more democratic approach to leadership towards the end of the study period.

The synchronization characteristics observed during the five rehearsal sessions in relation to the clear homophonic piece were paired with an overall horizontal tendency to tune to ET, avoiding pitch drift, and a vertical tendency to just intonation in the tuning of major thirds, and to ET in the tuning of minor thirds. These synchronization and tuning outcomes were also paired with a range of strategies adopted to solve tuning related issues, including tuning doubled notes, whole chords, specific melodic intervals, and balancing voices. Most of the changes were observed between the first two rehearsals, in terms of an improvement in the precision and consistency of synchronization in piece A and piece B, and consistency of tuning in piece A, alongside a tendency to tune horizontally slightly sharper than ET predictions.

This novel analysis of note timing and intonation is able to enrich the study of synchronization, revealing developmental aspects of synchronization and leadership in the context of performed tuning and in light of the verbal interactions and rehearsal strategies adopted.

Having described the changes in interpersonal synchronization resulting from altered visual contact in chapter *5*, and observed across rehearsals in the current chapter, this thesis goes on to analyse whether such patterns are reflected in the perception of the asynchronies resulting from the presence/absence of visual contact and degree of rehearsal, by listeners with different levels of expertise.

7 Perception of Singing Ensemble Synchronization

7.1 Introduction

The previous two empirical investigations reported in chapter *5* and chapter *6* have studied the roles of visual contact in interpersonal synchronization and the evolution of synchronization between singers during singing ensemble performances, as physically manifested in sound recordings. Results from these studies demonstrated that the precision of synchronization was better with visual contact between singers than without in singing duos (see chapter *5*), and improved between the first two rehearsals in a singing quintet rehearsing a homophonic piece (piece A) across five rehearsal sessions (see chapter *6*). It is now of interest to investigate whether these differences are perceivable by listeners with different levels of musical expertise (i.e., experts and non-experts) and types of musical expertise (i.e., singers and instrumental players; musicians performing or not in the study). This allows to consider the relevance of the results of the previous chapters in real-life performance situations during which members of the ensemble and the audience listen to the performed music. The current chapter investigates this aspect, addressing the following question:

- Do the differences in the asynchrony, physically measured in the singing ensemble performances earlier in the thesis, reflect the perception of synchronization by listeners with a variety of levels of musical experience?

As noted in section *2.4*, previous investigations show that asynchronies are not always perceived by listeners, and several elements might contribute to the detection of asynchronies, such as types of tones, and differences in frequency, duration, and sound intensity between two tones (Wallach, Newman, and Rosenzweig 1949; Hirsh 1959; Pastore, Harris, and Kaplan 1982; Bregman and Pinker 1978; Bregman 1990; Goebl 2003). The threshold of asynchrony perceptibility and temporal order discrimination is likely to increase with the complex onset behaviours of acoustic music events such as acoustic piano tones (Wallach, Newman, and Rosenzweig 1949; Goebl and Parncutt 2001; Goebl and Parncutt 2002; Goebl 2003). Research has also showed that the temporal order discrimination can change with experience of the listeners (Goebl 2003). On the other hand, research also found that listeners without specialized music training were sensitive to the degree of asynchrony when judging the level of togetherness in string quartet performances (Wing, Endo, Yates, et al. 2014), and the preference for trio jazz performances featuring asynchronies smaller than 19 ms was not related to the expertise of the listeners (i.e, musicians, non-musicians, and dancers) (Hofmann, Wesolowski, and Goebl 2017). However, the sensitivity to asynchrony from listeners with different levels of expertise, when judging ensemble singing, has not been fully investigated.

Although this perceptual study is explorative in nature, it was conjectured, more broadly, that listeners might not perceive differences in the synchronization if the asynchronies were smaller than 40 ms, which is considered to be the threshold for the perception of complex sounds, such as speech and music (Wallach, Newman, and Rosenzweig 1949). In other words, differences between asynchronies might not be perceived if the asynchronies in themselves are not perceivable.

7.2 Method

7.2.1 Participants

Ethical approval for the study (with reference D'Amario211117) was obtained from the Physical Sciences Ethics Committee (PSEC) at the University of York (UK). Thirty-three participants (15 females, 15 males, three non-binaries; $M = 25.2\,yo, SD = 4.2\,yo$) were recruited from the University of York (UK), comprising 10 undergraduate students, 16 postgraduate students, and seven doctoral researchers. Ten participants were from the Music Department, seven from the Psychology Department, and 16 from the Department of Electronic Engineering on a Music Technology related programme. Six students from the Department of Electronic Engineering were PhD students in the Audio Lab and had extensive experience in conducting perception studies not related to synchronization. As shown in **Table 7.1**, participants were classified based on their music training as non-experts and experts. The non-experts group comprised 20 university students with little or no music training. The experts group consisted of 13 advanced music students with at least four years of formal music training, and an average of 11.2 years formal music training ($SD = 4.2\,y$). The experts group comprised two sub-groups based on their performing experience: i) "performers in the study" group included five performers who performed the quintet recordings used as one subset of stimuli for the current study; and ii) the "other musicians" group comprised eight advanced music students. Based on their singing experience, the experts group was split into two more sub-groups: singers (seven) and instrumental players (eight). Three participants reported having some tinnitus, and two having absolute pitch. All participants received a nominal fee of £5.

Table 7.1 *Participant grouping based on music training, performing and singing experience.*

Classification	**Group (participants n)**
Music training	Experts (13); Non-experts (20)
Experience performing in the study	Experts (13): - Performers in the study (5); - Other musicians (8)
Advanced singing *vs* instrumental experience	Experts (13): - Singers (7); - Instrumental players (8)

7.2.2 Stimuli

Two sets of stereo recordings were presented to the listeners, comprising singing duo recordings (set A) selected from the study of visual contact reported in chapter *5*, and singing quintet performances (set B) selected from the study of the evolution of synchronization during rehearsal described in chapter *6*.

In chapter *5*, the analysis of synchronization was based on the full set of notes of a two-part piece relevant to the synchronization. In the current chapter, only the performance of the onset of the first note of the piece was used, as highlighted in **Figure 7.1**. This specific sound event was of great interest for perceptual analysis, since an effect of visual contact was found with the onset of the first note, and thus it was selected for the current investigation. Each onset token was approximately 1 s long, and was selected from the pool of singing duo performance tokens measured in chapter *5*. The entire pool comprised 576 absolute asynchronies temporally computed at the onset of the first note; half of the data set was performed in the presence of visual contact between singers and half without. As explained in detail in chapter 5 (see section *5.2.6*), these asynchronies were computed by subtracting the timing of one singer from that of the co-performer.

A subset of 48 tokens of 1 s (set A) was selected for the current perception study, to avoid fatiguing the listeners. To select the tokens, firstly, onsets of the entire data set were separately ranked for performances from smallest to largest asynchrony value with and without visual contact, resulting in two separate lists (i.e., one with and one without visual contact), each comprising 288 asynchronies. These two lists were then systematically sampled, choosing randomly the first asynchrony from each list and thereafter every 10th asynchrony. The fixed interval of 10 was arbitrarily decided to produce a total of 24 tokens per condition (i.e., with and without visual contact). This sampling procedure produced two lists of 24 snippets with an unbalanced number of tokens for each duo arising from picking every 10th asynchrony among the 288 ranked tokens performed by all 12 duos. Each list was therefore amended whereby the absolute values of a given (over-represented) duo selected through the sampling process would be replaced with the equivalent absolute values of a duo not represented in the list to create a final list that was balanced for duo as well as condition. This process assured that four tokens for each duo, two with and two without visual contact were selected.

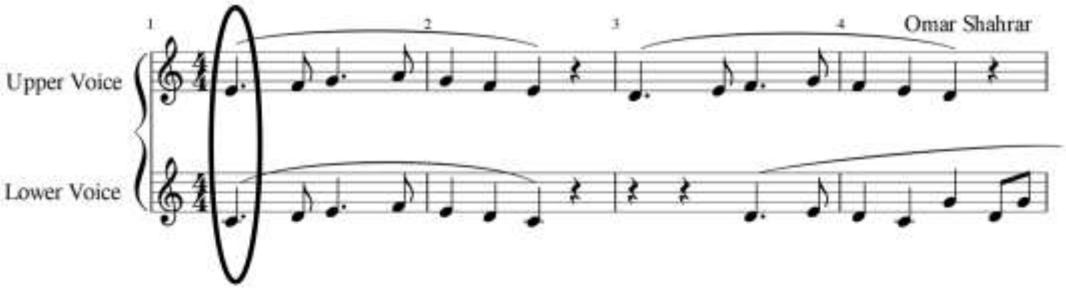



Figure 7.1 *The score of the first four bars of the two-part piece composed for previous studies, reported in chapter 3 and chapter 4, highlighting the first note (circled) that was used in the current perception study. The figure is from D'Amario, Daffern, and Bailes (2018b; 2018c) and is © the authors, licensed CC-BY.*

The other set of recordings, set B, comprised 10 quintet performances selected from the 30 recordings of quintet performances of the so named "piece A" (see **Figure 6.1**), used in the previous chapter to investigate the evolution of synchronization across five rehearsals. The full data set comprised recordings of the singing quintet who rehearsed piece A for 10 minutes in each rehearsal session, and performed three repetitions of the same piece before and after each rehearsal, as detailed in the previous chapter. This generated a total of six performances per rehearsal; each performance was approximately 50 s long. As detailed in chapter *6*, precision of synchronization was computed calculating the absolute asynchronies for each pair of singers at onsets and offsets of phonation, and note beginnings and endings within a legato phrase for each performance. From the original data set, a list of 10 recordings was selected featuring the two most representative performances of each rehearsal based on the following three-step process:

1. The average asynchronies were computed for each performance/onset/offset/beginning/ending, across pairs of singers. This step produced a mean synchronization value for each onset/offset/beginning/ending of each performance/session, resulting in 84 mean asynchrony values for each performance comprising two values for each note (i.e. onset/note ending at beginning of phonation, note beginning/ending within a legato phrase, and note beginning/offset at the ending of phonation).
2. Using the list generated from step one, the average asynchrony for each performance was computed, pooling together the 84 mean asynchrony values computed during the previous step. This process resulted in six mean asynchrony values (i.e., one per performance) for each rehearsal.
3. The grand means for each rehearsal were then calculated across the six mean values, and the two performances with the closest mean asynchrony values to the grand mean were selected for each rehearsal.

The stimuli were imported as .wav files into the Qualtrics online platform (www.qualtrics.com), and were presented using a desktop PC computer and headphones (Beyerdynamic DT990) at a level of 80 dB, which was measured and setup through a polystyrene dummy head with no pinnae, housing a microphone (NTi M4260) connected to a sound level meter (NTi XL2). The audio of the recordings was manually equalized across recordings, but not within the same recording. Therefore, all recordings were heard at approximately the same volume, and the expressive dynamics of the performance remained unchanged.

7.2.3 Design

As shown in **Table 7.2**, the study comprised two parts, namely Part A and Part B. Part A asked about the extent of synchronization in relation to the singing duo tokens (set A). Listeners were asked to answer the question: To what extent are the singers synchronized/together in time at the start of the following singing duo performances? Part B asked about the extent of synchronization in relation to the singing quintet performances (set B). Listeners were asked to answer the question: To what extent are the singers synchronized/together in time throughout the following quintet performances?

Parts A and B were presented in a counterbalanced order within each participant group (i.e. non-experts, performers in the study, and other musicians). The order of presenting the recordings was randomized within each part for each participant.

Table 7.2 *Research questions, stimuli, and questions posed to listeners in the study.*

Study part	Research questions	Stimuli	Listeners questions
Part A	RQ1: Do the differences in asynchronies resulting from altered visual contact reflect the perception of synchronization?	Set A: 48 singing duo tokens, ca 1 s long	Q1. To what extent are the singers synchronized/together in time at the start of the following singing duo performances?
Part B	RQ2: Do the differences in asynchronies measured across the five rehearsals reflect the perception of synchronization?	Set B: 10 singing quintet recordings, ca. 50 s long	Q2. To what extent are the singers synchronized/together in time throughout the following quintet performances?

7.2.4 Procedure

Participants were invited to an individual laboratory session that took place in a quiet room at the Department of Electronic Engineering of the University of York (UK) in February 2018. Participants first gave written, informed consent to take part in the study, and answered a questionnaire regarding their demographic background, music training, and singing experience. Then, the two parts of the study and the related questions and stimuli were presented. Listeners were asked to rate the extent to which the singers were synchronized on a sliding scale from zero to 100. Zero indicated that the singers were not at all synchronized/together in time; 100 meant they were fully synchronized/together in time. Listeners were able to play the recordings within the same set as many times as they wished, and they could change their rating through the sliding scale if needed. Once one of the two parts of the set was completed, listeners were not able to rectify their ratings or play the recordings of that set again.

Participants were asked to take a five-minute break between the two parts to reduce fatigue. An artificial break was introduced after the first half of the quintet recordings with a reminder to judge the extent to which singers were synchronized/together in time throughout the quintet performances. The full experiment lasted approximately 45 minutes.

7.2.5 Analysis

Stepwise multilevel linear-models of the response variable (i.e., perceived synchronization) were implemented to test the primary fixed effects of visual contact and rehearsal, and the random effects of participant. Listener groups based on expertise, performing or instrumental experience were also entered in the models as fixed effects nested within the primary fixed effects, as shown in **Table 7.3**. As argued in the previous two chapters, multilevel linear models were chosen because they increase the statistical reliability of the fixed effects analyses assessing the inter-participant variation (Gelman and Hill 2007). The models were implemented in R Studio (RStudio 2015) using the lme4 package.

Table 7.3 *Primary and nested fixed effects on the perceived synchronization, and the related recording sets used in the multilevel linear models.*

Model n	Primary fixed effects	Nested fixed effects	Recordings set
1	Visual contact	Expertise: experts (13), non-experts (20)	Set A: duo tokens
2	Visual contact	Instruments: singers (7), instrumental players (6)	Set A: duo tokens

3	Rehearsal number	Expertise: experts (13), non-experts (20)	Set B: quintet performances
4	Rehearsal number	Performing experience: performers in the study (5), other musicians (8), non-experts (20)	Set B: quintet performances
5	Performance asynchrony	Expertise: experts (13), non-experts (20)	Set A: duo tokens
6	Performance asynchrony	Instruments: singers (7), instrumental players (6)	Set A: duo tokens
7	Performance asynchrony	Expertise: experts (13), non-experts (20)	Set B: quintet performances
8	Performance asynchrony	Performing experience: performers in the study (5), other musicians (8), non-experts (20)	Set B: quintet performances

Perceived synchronization was the continuous response variable. Visual contact and rehearsal number were categorical explanatory variables with two (with and without visual contact) and five (rehearsal 1-5) categories, respectively. The nested fixed effect variables were also categorical variables. Performance asynchrony was a continuous explanatory variable.

The β coefficients on the predictor being considered are given below with reference to the specified base level of the factor, i.e. experts *versus* the base level non-experts, instrumental *versus* singing experience, and the "performers in the study" and "other musicians" groups *versus* the non-experts. As noted in the previous chapters, the β coefficients indicate that for each 1 unit increment in the predictor being considered, the effect of the given predictor changes by the amount specified by the β coefficient.

Prior to the analysis of the perception of synchronization, the role of visual contact and rehearsal number on the precision of performed synchronization was analysed, based on the newly created subset of duo tokens and quintet recordings respectively, which were selected from the original data sets reported in chapter *5* and chapter *6*. This was done to identify the main characteristics of the specific recordings used for the current perception study. As explained in section *7.2.2*, the sampling process used for the selection was done in such a way that the two subsets of data would be representative of the original, full data sets. Results from the subsets selected for the current study were expected to be similar but not exactly the same as those from the full data set. For this reason, it was important to analyse performed asynchronies of the data set of recordings selected for the current study. An independent t-test was conducted on the selected duo tokens to investigate whether there was a statistical difference between the means of synchronies measured in the presence and absence of visual contact. A regression model was implemented to test the primary fixed effect of rehearsal number on the precision of synchronization computed across the quintet performances selected for the current study. Note number and category (i.e., note beginnings and endings within a legato phrase, and onsets and offsets of phonation) were entered into the model as random variables.

Perceived ratings that fell outside three times the interquartile range (IQR) were automatically identified as extreme outliers through SPSS (IBM SPSS Statistics v. 24) and excluded from the analysis. A Bonferroni correction for multiple multilinear tests was implemented, and the alpha level was set at $p = 0.005$, based on a total of 9 models developed in the study.

7.3 Results

7.3.1 Performance Asynchrony

Prior to the analysis of the perception of synchronization, an initial test was conducted to investigate the effects of the presence/absence of visual contact on the precision of synchronization, based on the subset of duo token data selected for this study. An independent t-test demonstrated that synchronization in the subset of data was significantly better in the presence of visual contact between singers $(M = 57\ ms, SD = 44.9\ ms)$ than without it $[M = 95\ ms, SD = 72.5\ ms, t(46) = -2.2, p < 0.001]$, as shown in **Figure 7.2A**. The asynchronies measured with and without visual contact were on average 38 ms different; they were on average 57 ms with visual contact, and 95 ms without. As expected, this result corroborates the findings from the previous investigation reported based on the full data set of duo tokens showing that precision of synchronization was better with visual contact (see chapter *5*).

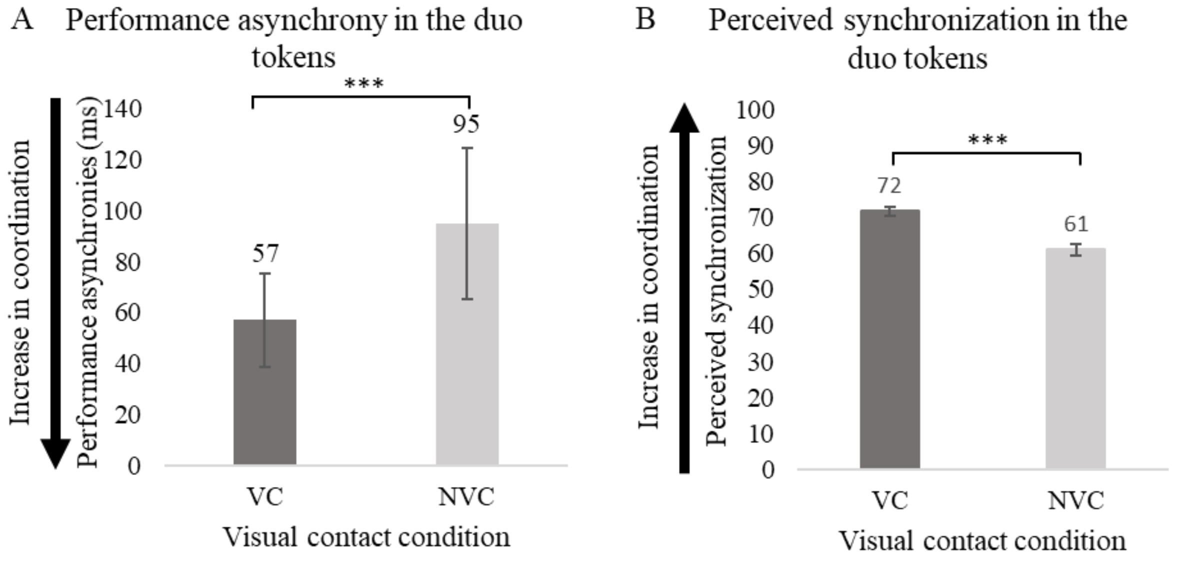


Figure 7.2 *Synchronization in the duo tokens selected for the current study and performed in presence and absence of visual contact between singers. Panel A displays the precision of the performed synchronization, and panel B ratings of perceived synchronization in a scale from 0-100, where 0 indicates singers were "not at all synchronized/together in time" and 100 "perfectly synchronized/together in time". Error bars represent 95 % CI of the mean.* $*** \; p < 0.001$.

Results from the regression analysis conducted on the quintet recordings selected for the current investigation demonstrate that, compared with rehearsal 1, precision of synchronization improved in rehearsal 2 $[\beta(-8.8), t(791.9) = -3.2, p < 0.01]$, in rehearsal 4 $[\beta(-9.1), t(791.9) = -3.4, p < 0.001]$, and rehearsal 5 $[\beta(-9.8), t(791.9) = -3.6, p < 0.001]$. *Post-hoc* tests show no significant changes across rehearsals 3-5, as shown in **Figure 7.3A**. The variance partition coefficient (VPC) among note number and category was 0.1084 and 0.1325, respectively; this indicates that 10.8 % and 13.3 % of the variability of the effect of rehearsal on performed synchrony can be attributed to note number and category, respectively. The asynchronies measured on this subset of data were on average larger than 49 ms, and on average up to 10 ms different across the five-rehearsal sessions. As expected, these results are very similar to those based on the full data set collected for chapter *6*.

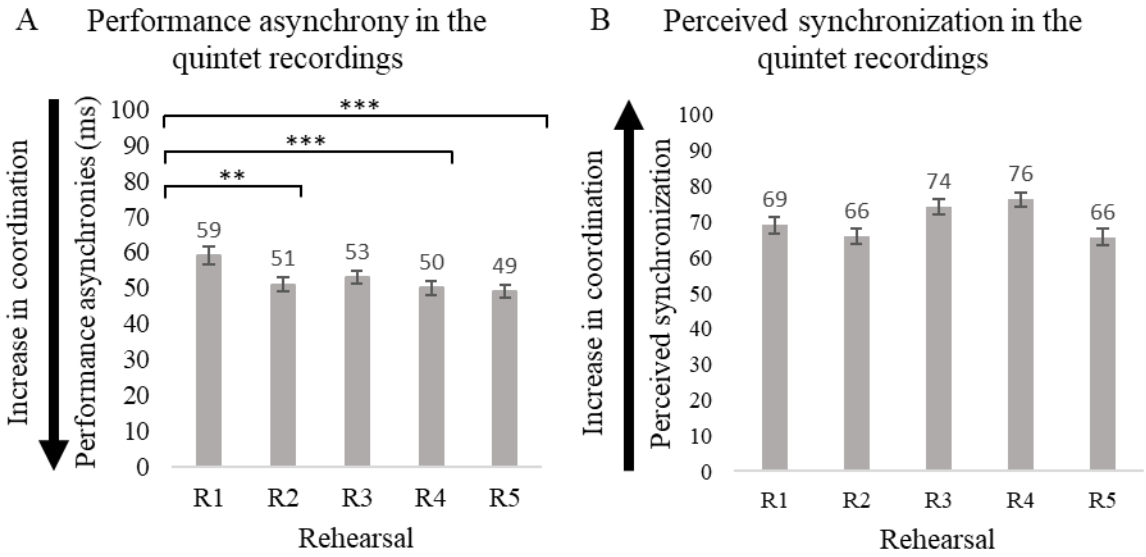


Figure 7.3 *Synchronization in the quintet recordings selected for the current study and performed across five rehearsal sessions. Panel A displays the precision of the performed synchronization, and panel B ratings of perceived synchronization in a scale from 0-100, where 0 indicates singers were "not at all synchronized/together in time" and 100 "perfectly synchronized/together in time". Error bars represent 95 % CI of the mean. p-values have been adjusted using the Holm method.* $** p < 0.01$; $***$ $p < 0.001$.

7.3.2 Perceived Synchronization

7.3.2.1 Visual Contact

Results from the multilevel linear modelling (model n 1 and n 2, as shown in **Table 7.3**) of perceived synchronization related to singing duo recordings show that visual contact between singers predicted ratings of perceived synchronization $[\beta(-8.8), t(1548) = -5.6, p < 0.001]$. Listeners rated the recordings performed in the presence of visual contact between singers $(M = 71.8\, ms, SD = 24.3\, ms)$ as being better synchronized than those performed without visual contact $(M = 61.0\, ms, SD = 29.2\, ms)$, as shown in **Figure 7.2B**.

Music training (i.e., experts *versus* non-experts) and instrumental experience (i.e., singers *versus* instrumental players) of respondents did not predict ratings of perceived synchronization based on the presence/absence of visual contact between singers, as shown in **Figure 7.4A** and **Figure 7.4B**. The variance partition coefficient (VPC) among participants was 0.186, which indicates that 18.6 % of the variability of the effect of visual contact on perceived synchrony can be attributed to the respondent category.

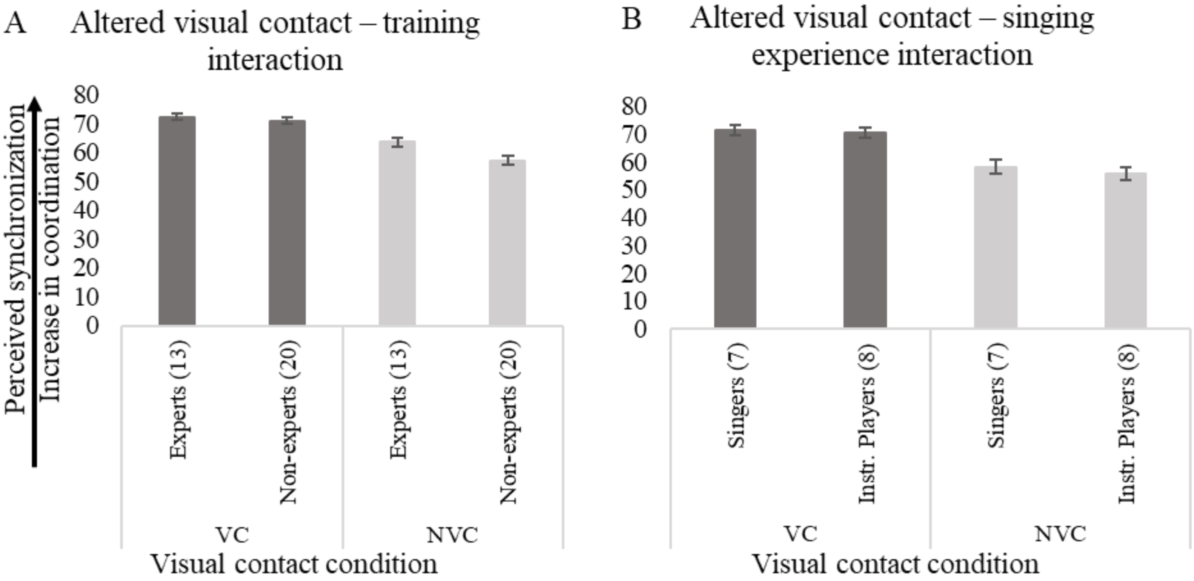


Figure 7.4 *Perception of synchronization by: A) interaction between presence/absence of visual contact and training level; and, B) interaction between visual contact condition and singing experience. The number of listeners in each subgroup (i.e., experts, non-experts, singers, and instrumental players) is given (n). Error bars represent 95 % CI of the mean.*

7.3.2.2 Rehearsal Number

The analysis of the effect of rehearsal in the singing quintet recordings (models n 3 and n 4, as shown in **Table 7.3**) demonstrates that rehearsal number (1-5) did not predict ratings of perceived synchronization (see **Figure 7.3B**). This is interesting in light of the synchronization measured in the recordings, highlighting an improvement in rehearsal 2, 4, and 5, compared with rehearsal 1, as show in **Figure 7.3A**.

This result did not differ according to the listeners' music training (i.e., experts *versus* non-experts) as shown in **Figure 7.5A**, or the performance experience of the listeners (i.e., non-experts, performers in the study, other musicians), as shown in **Figure 7.5B**. The VPC among participants was 0.42.

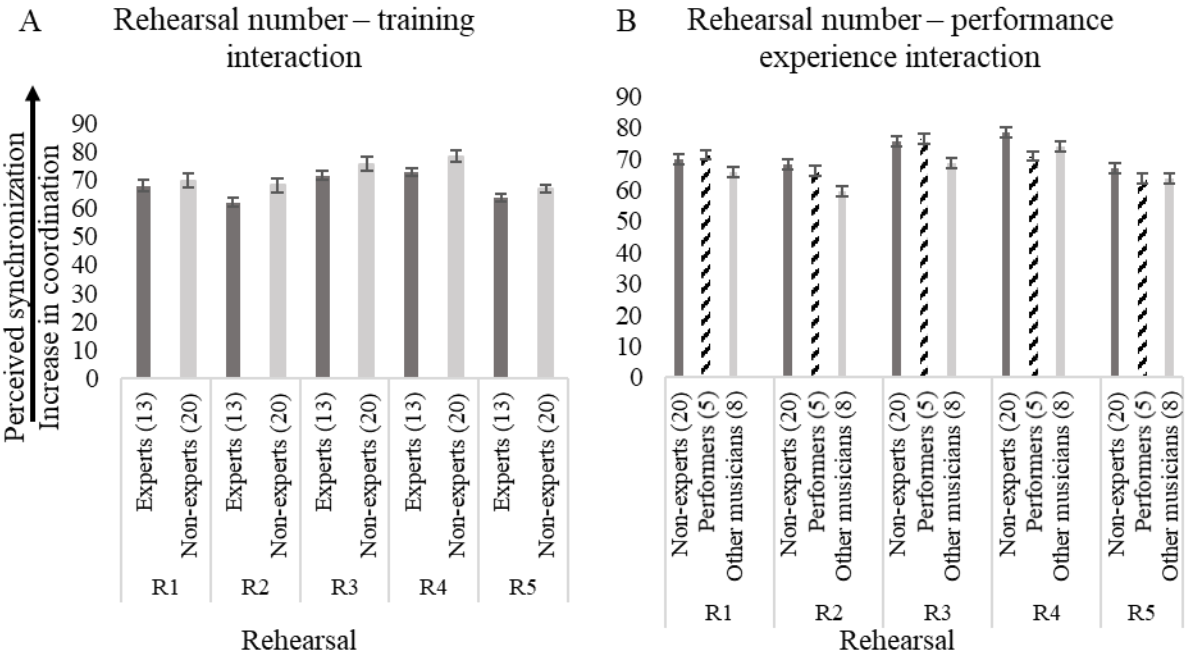


Figure 7.5 *Perceived synchronization based on degree of rehearsal (from rehearsal 1, R1, to rehearsal 5, R5) in relation to: A) the musical training of the listeners (i.e., non-experts, and experts); and B) the performance experience of the listeners (i.e., non-experts, performers in the study, and other musicians). The (*n*) represents the number of listeners in each subgroup (i.e., experts, non-experts, performers in the study, and other musicians). Error bars represent 95 % CI of the mean.*

Precision of the quintet's synchronization, indexed by absolute asynchronies, improved significantly in rehearsal 2, 4 and 5 compared with rehearsal 1, as shown in Fig 4B. *Post-hoc* tests were conducted to investigate whether listeners with different levels of training were able to perceive any differences between rehearsal 1-2, 1-4, and 1-5. This was not the case, irrespective of music training (i.e., experts *versus* non-experts) or performance experience of the listeners (i.e., non-experts, performers in the study, other musicians), as shown in **Figure 7.5A** and **Figure 7.5B**.

7.3.3 Relationship between Performance Asynchrony and Perceived Synchronization

Asynchrony measurements of the duo tokens and quintet recordings were entered stepwise in multilevel linear models (models n 5-8, see **Table 7.3**) to test whether the performance asynchronies predict the synchronization that listeners with different levels of expertise, training and performance experience can perceive.

Results show that there is a significant negative relationship between the performed asynchronies and the perceived synchronization of the duo tokens, as shown in **Figure 7.6A**. Specifically, 1 unit of reduced performance asynchrony (x-axis) is associated with an increase in perceived synchronization (y-axis) by 0.23 units $[\beta(-0.23), t(1573) = -22.2, p < 0.001]$. In other words, listeners perceived the smaller performance asynchronies as being better synchronized than the larger performance asynchronies. **Figure 7.6B** illustrates the relationship between performance asynchrony and perceived synchronization based on the music training of the listeners. The blue line representing the experts' ratings is steeper than the red one (non-experts' rating), suggesting that experts might tend to rate larger asynchronies as being less synchronized than non-experts. This finding illustrates a general trend which fails to reach significance. In addition, the singing *versus* instrumental experience is not a significant predictor of the perceived synchronization, as shown in **Figure 7.6C**.

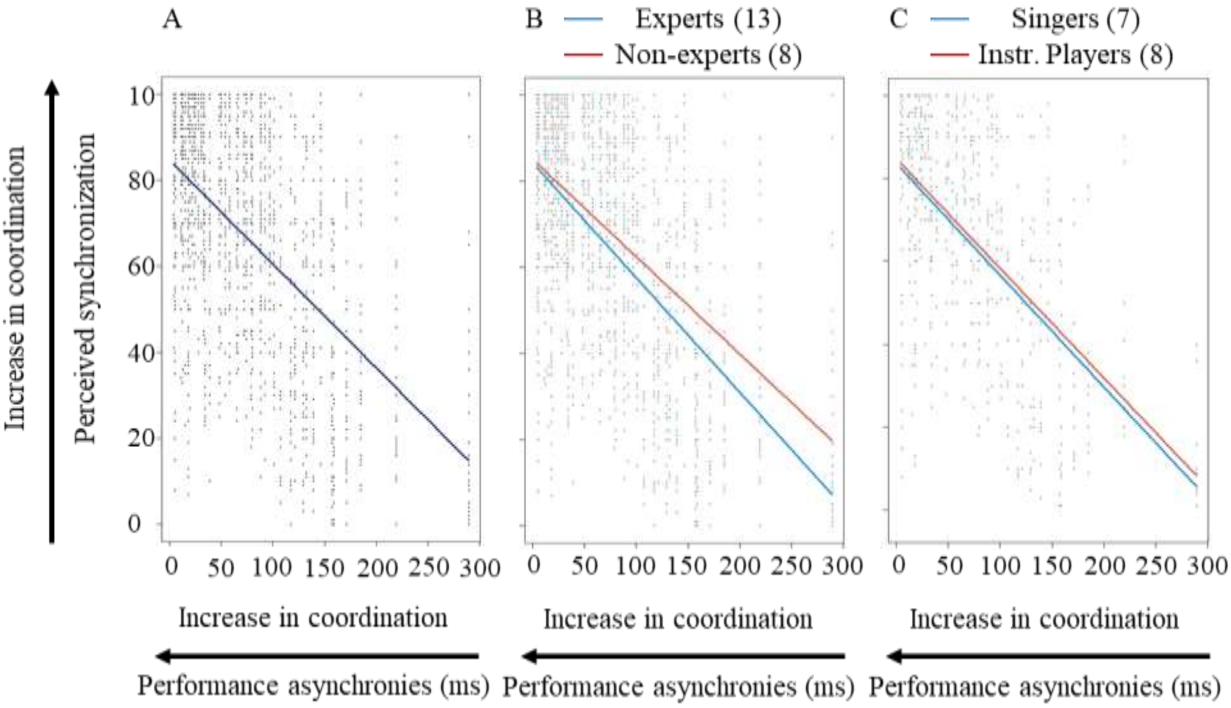


Figure 7.6 *Relationship between performance asynchrony and perceived synchronization in the duo tokens. Panel A displays the relationship between the performance asynchrony and the perceived synchronization; panel B shows the correlation between performance asynchrony and perceived synchronization in relation to the training of the listeners (i.e., experts and non-experts); and, panel C represents the correlation between performance asynchrony and perceived synchronization for singing and instrumental listeners.*

Results from the analysis of the quintet recordings demonstrate that the performance asynchronies measured across the five rehearsals did not predict ratings of perceived synchronization, irrespective of the music training, and types of performance experience of the listeners.

7.4 Discussion and Conclusions

This study investigates the perceptions of performance synchronization by listeners with various forms of music expertise, as this synchronization relates to differences resulting from altered visual contact and five rehearsal sessions. The relevance of the different degrees of physically measured synchronization in the previous two chapters was considered from a perceptual perspective in the current chapter, as this is a concern for performers and audience alike. Participants listened to two sets of recordings, which were taken from the two studies reported previously in the thesis, and which were representative of the findings from those studies. The sets of recordings comprised short duo tokens recorded with and without visual contact between singers, and full singing quintet performances recorded across five rehearsal sessions., and on average larger than 49 ms across the five rehearsals.

The duo tokens recorded in the presence of visual contact between singers were better synchronized than those without, as demonstrated by the analysis of the performance synchrony. The precision of synchronization, as quantified by the absolute asynchronies, was on average 57 ms in presence of visual contact between singers and 95 ms without. These differences in the performance synchrony were reflected in the perception of synchronization, irrespective of the music training and singing *vs* instrumental experience of the listeners. Listeners rated the tokens recorded with visual contact as better synchronized than those without visual contact. Specifically, 1 unit of reduced performance asynchrony was associated with an increase in perceived synchronization by 0.023 unit.

The precision of synchronization in the singing quintet recordings selected for the current study was on average larger than 49 ms across rehearsals, and improved across the five rehearsals. Synchronization was better during the second, fourth and fifth rehearsal compared with the first one. However, listeners were not able to perceive the significant differences in performance asynchrony that were measured across rehearsals, irrespective of the music training and performance experience of the listeners.

Listeners accurately perceived differences between performances sung with and without visual contact that differed on average by 38 ms, but they did not perceive differences between quintet performances that differed across rehearsals by only 10 ms on average. This suggests that listeners can perceive differences in asynchronies only above a certain threshold placed somewhere between 10 and 38 ms. These results complement the literature investigating the precedence effect, showing a threshold of asynchrony perception of around 40 ms for the perception of complex sounds, such as speech and music (Wallach, Newman, and Rosenzweig 1949). These findings also complement the more recent investigation conducted by Goebl and Parncutt (2001) into temporal order discrimination, reporting a threshold of 30 ms for acoustic piano tones. Altogether, the results suggest that the threshold for asynchrony perception is around 40 ms for complex sounds; the threshold for the temporal order discrimination is around 30 ms for acoustic piano tones; and, the threshold to detect differences in asynchronies is between 10 ms and 38 ms.

The lack of differences in the perception of asynchronies that were on average up to 10 ms different might also be related to the use of complex recordings comprising five singers performing simultaneously. For a given temporal difference, its perception might be related to the number of tones, becoming more difficult as the number of simultaneous tones increases.

Moreover, the length of the recordings might have also affected differences in the perception of such asynchronies. Pastore, Harris, and Kaplan (1982) found that the temporal order discrimination threshold was strictly related to the duration of the stimulus; it was as small as 5 ms if the duration of sinusoidal stimuli was as small as 10 ms, but the threshold was about 12 ms when the duration of the stimuli was 300 ms. Similarly, the perceptual judgement of differences in the performance asynchrony in the current study might change based on the duration of the stimuli, and it might increase when the duration of the stimuli increases. Further investigations in this area are needed to investigate the perceptual threshold of the differences in relation to the length of the stimuli.

In addition, results from the current study expand findings based on the perceptions of string quartet performances, showing that listeners without a specific musical training were sensitive to the degree of asynchrony between performers when judging the level of togetherness in string quartet playing (Wing, Endo, Yates, et al. 2014). This study also further expands studies on the temporal order discrimination, showing that non-experts were not able to discriminate the temporal order to pure, and complex tones (Goebl 2003). Altogether, these results suggest the musical training of the listeners might come into play for temporal order discrimination, but might not be relevant to the evaluation of the degree of synchronization of ensemble performances.

Some duo tokens had audible breaths at the beginning of the snippet, which might have affected the ratings of synchronization. It would be interesting in future research to investigate the extent to which audible breaths facilitate the perception of synchronization. This study focused on a harmonic major-third dyad, taken from the beginning of the first note of a duo, and a clearly homophonic five-part piece. For greater ecological validity, future investigation should involve the addition of timing complexities through rhythm and polyphony.

In conclusion, this study showed that differences in the precision of synchronization measured in singing duo tokens in relation to altered visual contact were reflected in the perceptions of synchronization irrespective of the music training and performance experience of the listeners. However, the smaller differences in performance asynchronies observed across rehearsals in quintet recordings were not reflected in the perceptions of synchronization, irrespective of the music training and performance experience of the listeners. This study provides a new context, from the perspective of perception, with which to consider the results of the previous two chapters investigating the roles of altered visual contact and degree of rehearsal in interpersonal synchronization measured in the singing performances.

8 Conclusions

Interpersonal synchronization in music ensembles is an essential aspect of music performance because it contributes to the expressiveness of ensemble performances, alongside other fundamental elements of music performance, such as rehearsals and tuning. Previous studies, mostly investigating single, instrumental performance sessions, suggest that the presence/absence of visual contact and leadership instruction between musicians might influence interpersonal synchronization (Goebl and Palmer 2009; Keller and Appel 2010; Kawase 2014; Bishop and Goebl 2015; Zamm, Pfordresher, and Palmer 2015). Research has also shown that ensemble playing is characterized by complex patterns of leader-follower relationships (Timmers, Endo, and Wing 2013; Timmers et al. 2014; Badino et al. 2014; Glowinski et al. 2012). Temporal coordination might evolve with practice, as rehearsal is a constant feature of music ensembles; and, listeners might be sensitive to the degree of asynchronies during ensemble performances (Wing, Endo, Yates, et al. 2014). However, the impact of visual contact and leadership instruction on interpersonal synchronization between musicians during ensemble performances, and the evolution of synchronization and leader-follower relationship across rehearsals in singing ensembles have not yet been fully investigated. In addition, a robust method to investigate synchronization between singers during *a cappella* singing performances has not been developed.

This thesis first implemented and tested a novel algorithm to analyse synchronization in the context of ensemble singing. Then, through a number of empirical investigations based on the newly developed algorithm, the thesis has investigated: i) the roles of visual contact and leadership instructions in synchronization in singing duo ensembles; ii) the developmental aspects of synchronization and leadership across rehearsals in relation to the complexity of the piece being rehearsed in a singing quintet; and, iii) whether differences in the synchronization resulting from altered visual contact and degree of rehearsal are reflected in the perception of synchronization. In addition, the analysis of the evolution of synchronization across rehearsals has been conducted in parallel with the analysis of tuning outcomes and the verbal interactions between singers during rehearsal, to contextualise the findings related to synchronization within the wider scope of expressive performance behaviours.

The overall research question was formulated as follows: Do visual contact, leader-follower relationships, and rehearsal influence interpersonal synchronization during singing ensemble performances? Although this research was explorative in nature, it was hypothesised that *interpersonal synchronization is not affected by visual contact between singers during the performance of a two-part piece with regular metre, improves with practice across rehearsals, and is associated with a mutual adaptation of leader-follower roles*.

The novel algorithm, termed TIMEX, developed in this thesis to analyse synchronization in the context of ensemble singing, extracts onsets and offsets of phonation as well as note beginnings and endings within a legato phrase; and, it is based on the combined application of electrolaryngography and acoustic analysis, allowing fundamental frequency evaluation of the individual voices. The new algorithm, tested on a set of singing duo performances collected for the study of the role of visual contact and leadership instruction in synchronization, proved to be a *state-of-the-art* algorithm for timing detection in ensemble singing. The algorithm allowed consistency and repeatability of the extraction of timing information for approximately 60 % of the data collected and analysed, thus solving many of the issues regarding the subjectivity and laboriousness of the typical manual detection, especially in the context of ensemble singing (Devaney 2011).

This algorithm was then applied throughout the investigations reported in this thesis, to investigate interpersonal synchronization between singers during singing ensemble performances. The analysis of synchronization between singers has shown that the precision of synchronization was on average 59 ms in the singing duos and 63 ms in the singing quintet as quantified by the absolute asynchronies. The consistency of synchronization was on average 67 ms in the duo performances, but was better (on average 39 ms) across rehearsals in the quintet recordings, as measured by the standard deviation of absolute asynchronies. A better consistency across rehearsals in the singing quintet compared with the individual performance sessions of the singing duos is an expected result, in that as the singers practise, they improve their interpersonal synchronization. The tendency to lead was on average -4.06 ms in the duos, and 1.5 ms across rehearsals, as quantified by the median signed asynchronies. Taken together, these main synchronization characteristics are in line with the literature analysing synchronization during naturalistic ensemble performance and controlled experiments, reporting typical mean absolute and SD asynchronies to be between 30 and 50 ms, and mean signed asynchronies mostly close to 0 ms (Keller 2014; Goebl and Palmer 2009; Timmers et al. 2014; Bishop and Goebl 2015; Zamm, Pfordresher, and Palmer 2015).

This thesis has then analysed the role of visual contact and has shown that the above synchronization characteristics were affected by the presence and absence of visual contact between duo singers during the repeated performances of a mostly homophonic two-part piece. The precision and consistency of synchronization were better with visual contact than without throughout the repeated performances. The thorough analysis of the impact of these variables conducted also at the lower levels [i.e., at crucial points of the performance, and in relation to the time categories considered (onsets and offsets of phonation, and note beginnings and endings within a legato phrase)] has provided a novel approach to the study of synchronization, revealing insights into the nature of these effects.

The influences of visual contact on the precision and consistency of synchronization were mostly seen at the beginning of phonation of the piece, demonstrating that musicians performing in the absence of visual contact were able to compensate soon after. With visual contact, the precision and consistency were found to differ according to the time category considered, and synchronization being less precise and consistent at the beginning of the piece, compared to other simultaneous entry points. The absence of visual contact accentuates these differences: synchronization at the onset was found to be even less precise and consistent compared with that computed at other simultaneous entry points. The combined analysis of the precision of synchronization as measured in the sound produced at the beginning of the piece, in parallel to its perception, has further advanced our understanding of the role of visual contact. Results showed that the differences in the performed synchronization resulting from altered visual contact were reflected in differences in the perception of synchronization. The perception of these differences did not depend on the different levels of musical expertise of the listeners, implying the importance of synchronization transcending the listeners’ musical expertise. Finally, the presence/absence of visual contact also affected the tendency to precede/lag a co-performer. In the presence of visual contact, the designated leader was found to be ahead by a small, but significant amount of time; but without visual contact, the designated leader showed overall a stronger tendency to precede the designated follower. Similarly to the precision and consistency of synchronization, this effect was found to be associated with the onset of note 1 of the piece.

Another aspect of interest in this thesis was the roles of leadership, conceptualized as tendency to precede/lag the co-performer(s) during ensemble performances. Results have shown that instructing the singers to act as leader or follower during the singing duo performances did not influence interpersonal synchronization, but it affected the tendency to precede/lag a co-performer. The results regarding the precision and tendency to lead/lag a co-performer changed with the time category considered. When the upper voice was instructed to follow, the precision was better at onsets of phonation than note beginnings within the legato phrases; these differences were amplified when the upper voice was instructed to lead. The tendency to precede/lag was larger at onset of phonation than offset of phonation and note beginning/ending within the legato phrase. Instructing the upper voice to lead emphasised these differences with the upper voice showing a stronger tendency to precede the co-performer at the onset of phonation, compared with when instructed to follow.

The above results regarding the roles of visual contact and leadership instruction in interpersonal synchronization during singing duo performances, were consistent across singers, suggesting that these results might typify this type of ensemble when performing the two-part piece, mostly homophonic, composed for this study.

A further aim of this thesis was to investigate the evolution of synchronization and leader-follower relationships emerging spontaneously (i.e., without leadership instruction) across rehearsals. Results have shown that interpersonal synchronization and leader-follower relationships changed in a newly formed, semi-professional *a cappella* singing quintet across five rehearsals during a four-month study period. The characteristics of these changes were related to the complexity of the piece being practised. The precision of synchronization improved between the first two rehearsals in the less complex piece, piece A, and consistently across the five rehearsals in the most complex piece, piece B. Similarly, changes in the consistency of synchronization were related to the piece practised: consistency did not change in piece A, but increased in the first two rehearsals in piece B. There were no explicit implementations of rehearsal strategies to solve timing issues to reflect these patterns. Changes in the precision of synchronization across rehearsals in the less complex piece, piece A, were not reflected in the perceptions of synchronization, irrespective of the levels of the listeners' musical expertise. The tendency for members of the ensemble to precede/lag the others changed across the five rehearsals: it was not equally shared between singers during early rehearsals, but it became equally shared by the last rehearsal session.

Finally, through the novel application of a mixed-method approach, involving quantitative and qualitative investigations of both temporal coordination and tuning behaviours, this thesis has revealed a complex picture of synchronization in the less complex piece, piece A, understood within the context of the wider quest for expressiveness in ensemble performance. Synchronisation was analysed in parallel to the tuning outcomes, and in light of the verbal interactions and rehearsal strategies adopted. The above synchronization characteristics observed across rehearsals, in fact, were paired with a consistent and repetitive tendency to tune horizontally to equal temperament (ET), avoiding pitch drift, in the presence of examples of just intonation occurring frequently in the vertical tuning of major thirds, and instances of ET tuning in the minor thirds.

Specifically, each singer performed closer to ET, avoiding pitch drift based on just intonation predictions, during and between rehearsals and across repetitions. Deviation from equal temperament was flatter towards the end of the four-month study period compared with the initial rehearsal, and was more consistent and narrower in the third rehearsal. The ensemble tuned the major thirds slightly closer to just intonation, and the minor thirds closer to equal temperament, and these results were consistent within and between rehearsals, but changed based on chord number and pair of singers. The change based on chord number and pair of singers suggests that these results might be specific to the piece rehearsed and might typify this singing quintet. Furthermore, these tuning outcomes were also associated with a range of strategies adopted to solve tuning related problems, such as tuning doubled notes, specific chords and melodic intervals, and balancing voices.

In addition to the emerging avenues for further research strictly connected to the specific empirical investigations reported in the relevant chapters, a number of other pathways are suggested for the future, which apply to the whole thesis.

Firstly, this research concerned semi-professional singing duos and a semi-professional singing quintet. To understand whether the above results typify the size and type of these ensembles, it will be necessary for future studies to build a corpus of research which will gradually examine the consistency of the findings across performances of different types and sizes of ensemble. Some of the above synchronization patterns might change according to different levels of music expertise of the musicians; interpersonal synchronization might be less precise and consistent with amateur ensembles rather than semi-professional or professional ensembles. The results of this thesis should be tested with amateurs and/or professional ensembles to advance our understanding of the evolution of synchronization with increasing musical expertise of the musicians.

In this research, the tendency to precede or lag timing is considered an indicator of leader-follower relationships between the singers, which is common in this field of research (Goebl and Palmer 2009; Keller and Appel 2010; Timmers et al. 2014; Timmers, Endo, and Wing 2013; Palmer et al. 2013; Zamm, Pfordresher, and Palmer 2015). Nevertheless, the fact that one of the musicians might tend to anticipate or lag another is not a comprehensive perspective on leadership, which can be viewed more in terms of social roles (King 2006; Carnicer, Garrido, and Requena 2015; Lim 2014; Page-Shipp, Joseph, and van Niekerk 2018) rather than in terms of performance timing. The study of leader-follower relationships based on the combined analysis of synchronization during ensemble performances and patterns of social interactions emerging during rehearsals might shed more light on our understanding of leader-follower relationships in singing ensembles.

Furthermore, the above results might be strictly related to the characteristics of the stimuli used for this research (i.e., a mostly homophonic two-part piece; two mostly homophonic five-part pieces with different melodic contours and harmonic structure from each other, without a clear separation of leader-follower roles; a major third interval selected from the singing duo recordings for the perception study; and, the vowel selected for the performance). Future investigations involving different musical excerpts with lyrics and greater melodic, tonal, rhythmical, and leadership complexity should be conducted, to evaluate the replicability of the above findings in more musically complex pieces, verifying the generalisability of the results.

It would also be interesting to replicate this research with different genres, such as jazz ensembles. Studies conducted by Mark Doffman and analysing mostly single performance sessions have shown dynamic temporal variations between pairs of musicians of tight and loose coordination (Doffman 2008; Doffman 2009; Doffman 2013). The current investigation revealed that the effect of visual contact was mostly seen at the beginning of the piece, and precision and consistency of synchronization improved across rehearsals in a Classical singing quintet, depending on the piece being practised. The jazz tradition might elicit different results.

Variation of these findings between cultures might also be an important aspect of the role of visual contact and of the evolution of synchronization and leadership across rehearsals. Interpersonal synchronization in cultures with a less systematic approach to rehearsal than the Western Classical tradition, such as African oral traditions, or a clear division of group roles, as in Transylvanian music traditions, might elicit different outcomes during and across rehearsals.

This research contributes to the study of interpersonal synchronization between musicians during ensemble performances by:

- Investigating the main characteristics of synchronization during repeated performances, comprising the precision and consistency of synchronization and the tendency to precede/lag a co-performer, in 12 singing duos and in one *a cappella* singing quintet.
- Understanding the role of visual contact and leadership instruction in singing duo performances.
- Observing the evolution of synchronization and leader-follower relationships across rehearsals during a four-month study period.
- Contextualising the findings of synchronization within the wider scope of expressive performance goals, analysing the developmental aspects of synchronization and leadership in parallel with tuning behaviours and in light of the verbal interactions between singers and the rehearsal strategies adopted.
- Underscoring the salience of synchronization, which transcends the musical expertise of the listeners.
- Providing a novel and successful algorithm for timing detection in ensemble singing that addresses much of the subjectivity and laboriousness of typical manual annotation.

In addition, the findings of this research might point to ways in which rehearsals could be optimised to improve synchronization in singing ensembles. Singers and directors can be better informed of the factors that might influence synchronization (i.e., visual contact, leadership, rehearsal, and the rhythmical complexity of the piece) and on the rapidity at which the precision and consistency of synchronization at which the singing quintet seems to settle down. Musicians and music teachers can be also informed on the focus on specific areas of difficulty in certain performance conditions, such as first note onsets when performers are not able to see each other. Choir directors and coaches have an evidence base and context from which their rehearsal strategies and performance goals can develop. The results of this study are also of interest for psychology research, aimed at clarifying the psychological processes that characterize interpersonal communication, non-verbal communication, and social interaction.

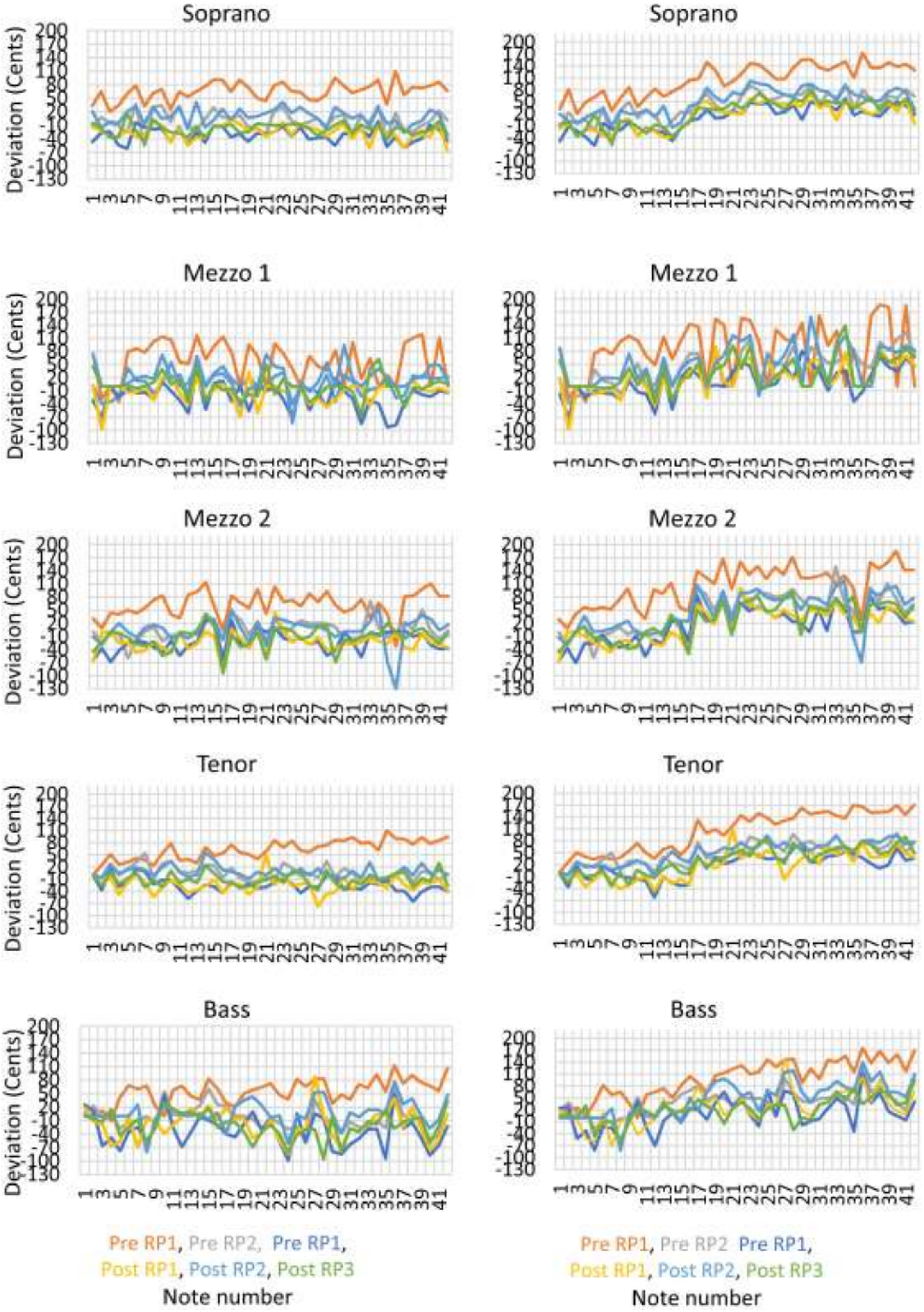
The novel algorithm presented in this thesis, which detects onsets/offsets of phonation, and note beginnings/endings within a legato phrase, and which was developed for the analysis of synchronization in ensemble singing, might be useful for a range of aspects of music performance analysis and audio signal processing, such as music information retrieval and transcription applications.

To conclude, this thesis has shown that interpersonal synchronization between singers during singing ensemble performances in the Western Classical tradition represents a complex phenomenon, depending on the presence/absence of visual contact and degree of rehearsal, and characterized by a complex pattern of leader-follower relationships. Future investigations should investigate the generalizability of these findings, replicating this research on the basis of different music pieces, ensembles, genres, and cultures.

This research has paved the way for future investigations to consider the study of synchronization within the wide scope of expressive ensemble performance, analysing simultaneously several expressive elements of ensemble playing. Future technologies should enable rapid progress in the multi-layered studies that can build on this research by analysing different levels of synchronization simultaneously, from neural processes, to physiological mechanisms, to body movements, and sound production, to reveal the complex mechanisms underpinning synchronization in music ensembles.

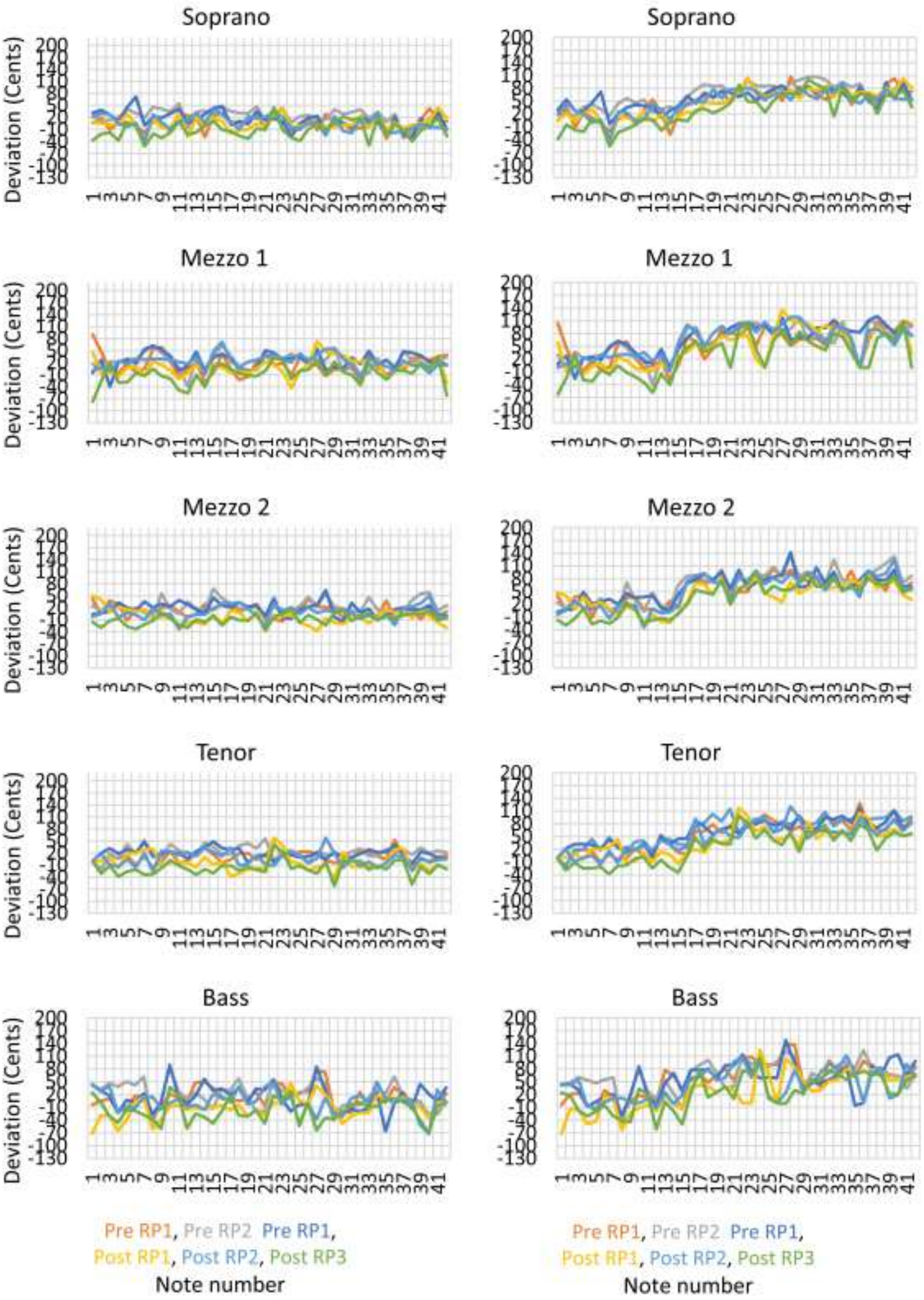
Appendix A. Measured Tuning Deviation across Rehearsal 1

Tuning deviation from expected equal temperament (on the left-hand side) and predicted just intonation (right-hand side) is computed for each singer and repetition.



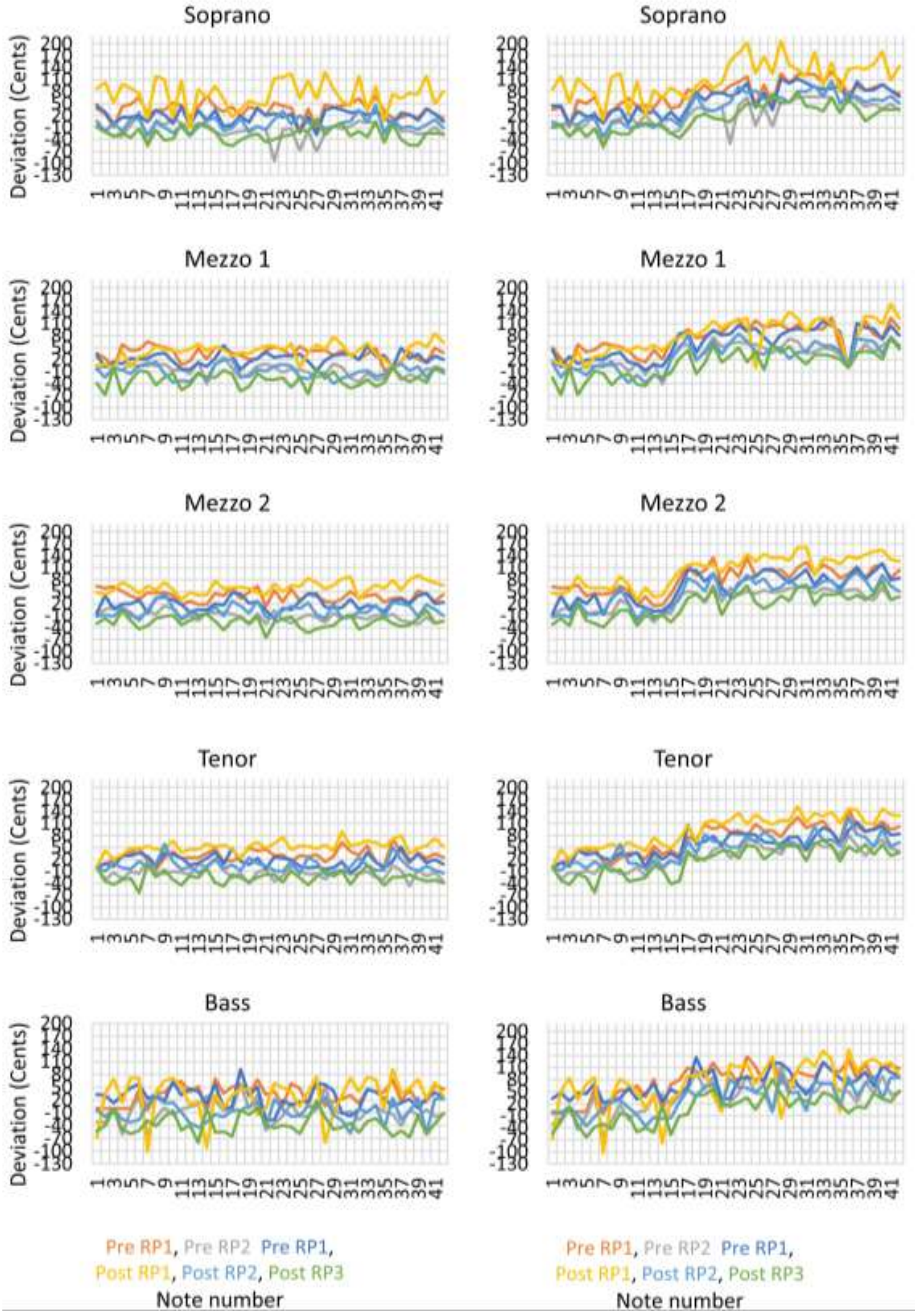
Appendix B. Measured Tuning Deviation across Rehearsal 2

Tuning deviation from expected equal temperament (on the left-hand side) and predicted just intonation (right-hand side) is computed for each singer and repetition.



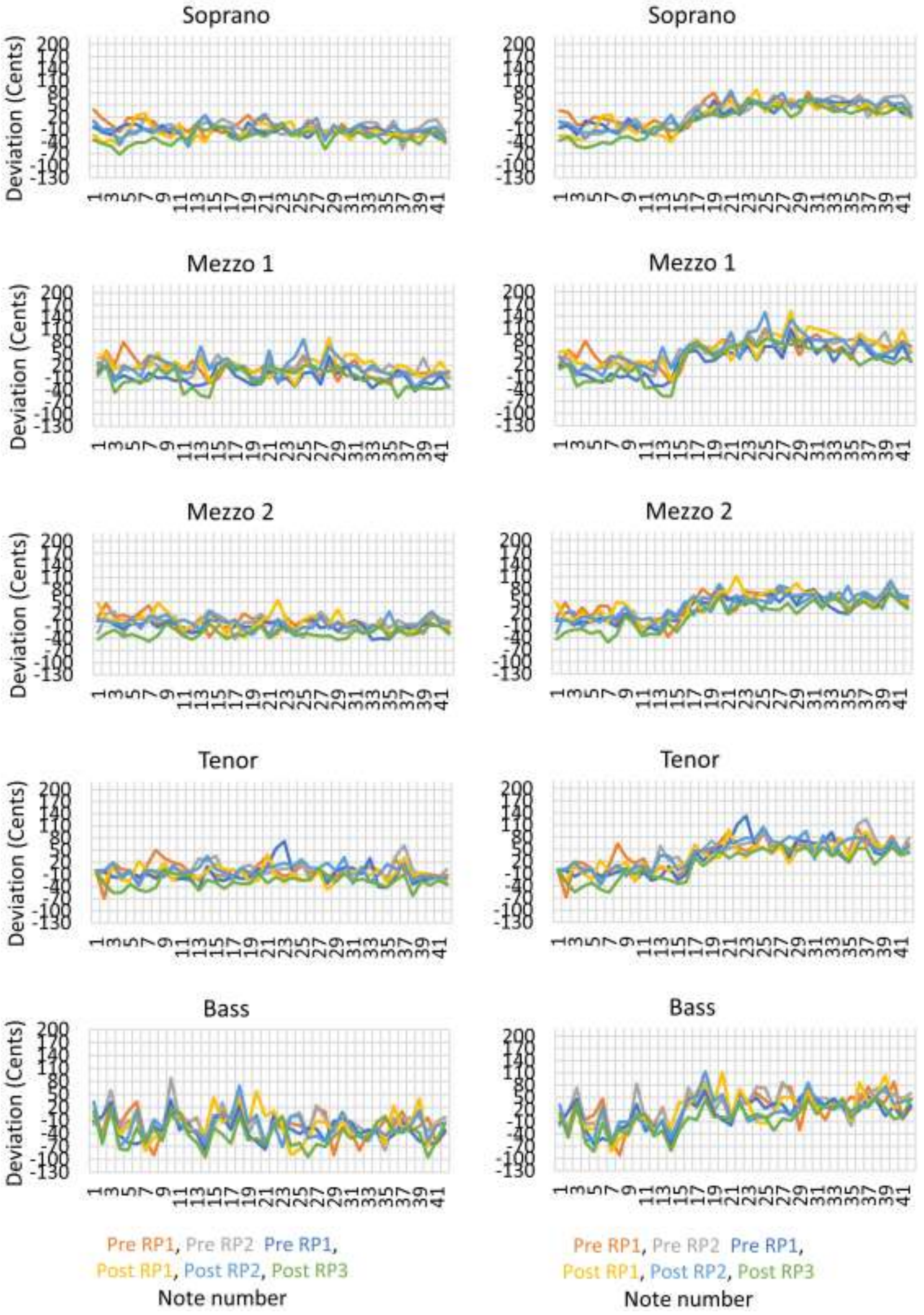
Appendix C. Measured Tuning Deviation across Rehearsal 3

Tuning deviation from expected equal temperament (on the left-hand side) and predicted just intonation (right-hand side) is computed for each singer and repetition.



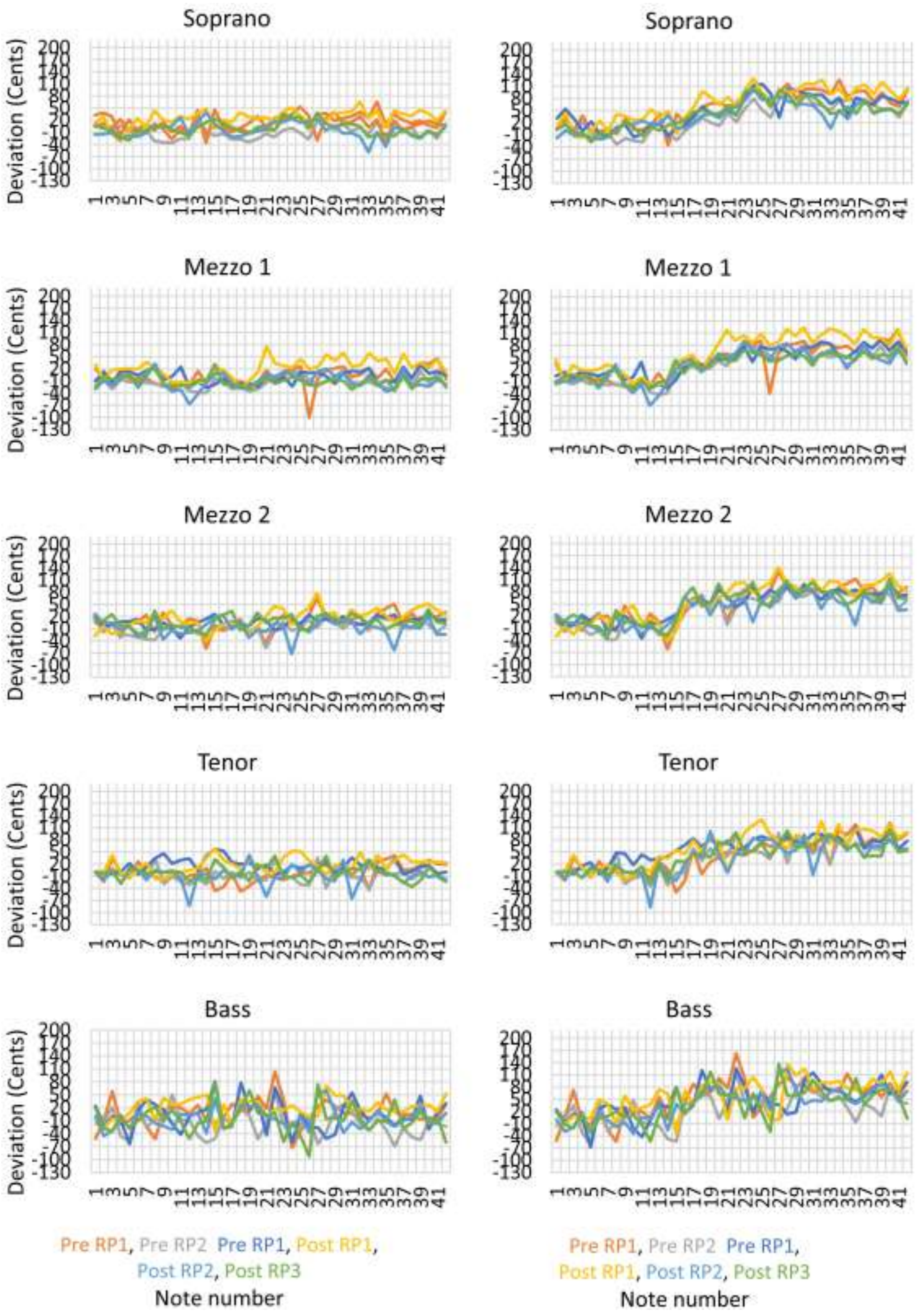
Appendix D. Measured Tuning Deviation across Rehearsal 4

Tuning deviation from expected equal temperament (on the left-hand side) and predicted just intonation (right-hand side) is computed for each singer and repetition.



Appendix E. Measured Tuning Deviation across Rehearsal 5

Tuning deviation from expected equal temperament (on the left-hand side) and predicted just intonation (right-hand side) is computed for each singer and repetition.



Appendix F. List of Symbols and Acronyms

Amp	amplitude
β	beta coefficient
BPM	beats per minute
CP	vocal fold closed phase
CQ	vocal fold contact quotient
EEG	electroencephalography
EGG	electroglottography
EMG	electromyography
ET	equal temperament
F	f-measure
f_o	fundamental frequency
fMRI	functional magnetic resonance imaging
FN	False negative
FP	false positive
Lx	electrolaryngography
NB	note beginning
NE	note ending
NVC	without visual contact
OF	offset
ON	onset
OP	vocal fold open phase

P (1-5)	position (1-5)
Post-	post-rehearsal
Pr	precision
Pre-	pre-rehearsal
R (1-5)	rehearsal (1-5)
Re	recall
RP	repeated performance
SATB	soprano-alto-tenor-bass
T_0	fundamental period
UpperVoiceF	upper voice follows
UpperVoiceL	upper voice leads
VC	with visual contact
VRP	variance partition coefficient

References

Abdallah, Samer, and Mark Plumbley. 2003. “Unsupervised Onset Detection: A Probalistic Approach Using ICA and a Hidden Markov Classifier.” In *Proceedings of Cambridge Music Processing Colloquium*.

Åkerlund, Leif, Patricia Gramming, and Johan Sundberg. 1992. “Phonetogram and Averages of Sound Pressure Levels and Fundamental Frequencies of Speech: Comparison between Female Singers and Nonsingers.” *Journal of Voice* 6 (1): 55–63. doi:10.1016/S0892-1997(05)80009-8.

Askenfelt, Anders, Jan Gauffin, and Johan Sundberg. 1980. “A Comparison of Contact Microphone and Electroglottograph for the Measurement of Vocal Fundamental Frequency.” *Journal of Speech and Hearing Research* 23 (2): 258–73.

Babiloni, Claudio, Fabrizio Vecchio, Francesco Infarinato, Paola Buffo, Nicola Marzano, Danilo Spada, Simone Rossi, Ivo Bruni, Paolo M. Rossini, and Daniela Perani. 2011. “Simultaneous Recording of Electroencephalographic Data in Musicians Playing in Ensemble.” *Cortex* 47 (9): 1082–90. http://dx.doi.org/10.1016/j.cortex.2011.05.006.

Badino, Leonardo, Alessandro D’Ausilio, Donald Glowinski, Antonio Camurri, and Luciano Fadiga. 2014. “Sensorimotor Communication in Professional Quartets.” *Neuropsychologia* 55: 98–104. http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.012.

Barlow, Chris, Jeannette Lovetri, and David M. Howard. 2007. “Voice Source and Acoustic Measures of Girls Singing ‘Classical’ and ‘Contemporary Commercial’ Styles’.” *Proceedings of the International Symposium on Performance Science, Casa Da Musica, Porto, Portugal*, 195–200. http://www.legacyweb.rcm.ac.uk/cache/fl0020220.pdf.

Bartholomew, Wilmer. 1934. “A Physical Definition of ‘Good Voice-Quality’ in the Male Voice.” *The Journal of the Acoustical Society of America* 6 (1): 25–33. https://doi-org.libproxy.york.ac.uk/10.1121/1.1915685.

Bello, Juan P., Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. 2005. “A Tutorial on Onset Detection in Music Signals.” *IEEE Transactions on Speech and Audio Processing* 13 (5): 1035–46. doi:10.1109/TSA.2005.851998.

Bello, Juan P., Chris Duxbury, Mike Davies, and Mark B. Sandler. 2004. “On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain.” *IEEE Signal Processing Letters* 11 (6): 553–56. doi:10.1109/LSP.2004.827951.

Bello, Juan P., and Mark B. Sandler. 2003. “Phase-Based Note Onset Detection for Music Signals.” In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-03), Hong Kong*, 49–52. doi:10.1109/ASPAA.2003.1285811.

Biasutti, Michele, Eleonora Concina, David Wasley, and Aaron Williamon. 2016. “Music Regulators in Two String Quartets: A Comparison of Communicative Behaviors between Low- and High-Stress Performance Conditions.” *Frontiers in Psychology* 7 (August): 1229. https://doi.org/10.3389/fpsyg.2016.01229.

Bishop, Laura, and Werner Goebl. 2014. “Context-Specific Effects of Musical Expertise on Audiovisual Integration.” *Frontiers in Psychology* 5 (October): 1123. https://doi.org/10.3389/fpsyg.2014.01123.

———. 2015. “When They Listen and When They Watch: Pianists’ Use of Nonverbal Audio and Visual Cues during Duet Performance.” *Musicae Scientiae* 19 (1): 84–110. https://doi.org/10.1177/1029864915570355.

———. 2017. “Beating Time: How Ensemble Musicians’ Cueing Gestures Communicate Beat Position and Tempo.” *Psychology of Music* 46 (1): 84–106. https://doi.org/10.1177/0305735617702971.

———. 2018. “Communication for Coordination: Gesture Kinematics and Conventionality Affect Synchronization Success in Piano Duos.” *Psychological Research* 82 (6): 1177–94. https://doi.org/10.1007/s00426-017-0893-3.

Blank, Marilyn, and Jane W. Davidson. 2007. “An Exploration of the Effects of Musical and Social Factors in Piano Duo Collaborations.” *Psychology of Music* 35 (2): 231–48. https://doi.org/10.1177/0305735607070306.

Blum, David. 1987. *The Art of Quartet Playing: The Guarneri Quartet*. New York: Cornell University Press.

Boersma, Paul. 2001. “Praat, a System for Doing Phonetics by Computer.” *Glot International*

5 (9/10): 341–47.

Boersma, Paul, and David Weenink. 2013. “Praat: Doing Phonetics by Computer.” http://www.praat.org/.

Bonini-Baraldi, Filippo, Emmanuel Bigand, and Thierry Pozzo. 2015. “Measuring Aksak Rhythm and Synchronization in Transylvanian Village Music by Using Motion Capture.” *Empirical Musicology Review* 10 (4): 265–91. http://dx.doi.org/10.18061/emr.v10i4.4891.

Bonsignore, Maria R., Salvatore Romano, Oreste Marrone, and Giuseppe Insalaco. 1995. “Respiratory Sinus Arrhythmia during Obstructive Sleep Apneas in Humans.” *Journal of Sleep Research* 4 (Suppl 1): 68–70. https://doi.org/10.1111/j.1365-2869.1995.tb00190.x.

Bregman, Albert S. 1990. *Auditory Scene Analysis. The Perceptual Organization of Sound*. Cambridge, Massachusetts: The MIT Press.

Bregman, Albert S., and Steven Pinker. 1978. “Auditory Streaming and the Building of Timbre.” *Canadian Journal of Psychology* 32: 19–31.

Brereton, Judith S. 2014. “Singing in Space (s): Singing Performance in Real and Virtual Acoustic Environments — Singers’ Evaluation, Performance Analysis and Listeners’ Perception.” University of York.

Brown, William S., Richard J. Morris, Douglas M. Hicks, and Elizabeth Howell. 1993. “Phonational Profiles of Female Professional Singers and Nonsingers.” *Journal of Voice* 7 (3): 219–26. https://doi.org/10.1016/S0892-1997(05)80330-3.

Cannam, Chris, Christian Landone, and Mark B. Sandler. 2010. “Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files.” In *Proceedings of the ACM Multimedia 2010 International Conference, October 2010, Firenze, Italy*, 1467–68.

Carnicer, Josep G., Diego C. Garrido, and Salvador O. Requena. 2015. “Music and Leadership: The Role of the Conductor.” *International Journal of Music and Performing Arts* 3 (1): 84–88. http://dx.doi.org/10.15640/ijmpa.v3n1a8.

Castellano, Ginevra, Marcello Mortillaro, Antonio Camurri, Gualtiero Volpe, and Klaus Scherer. 2008. “Automated Analysis of Body Movement in Emotionally Expressive Piano

Performances.” *Music Perception: An Interdisciplinary Journal* 26 (2): 103–19. https://doi.org/10.1525/mp.2008.26.2.103.

Chaffin, Roger, and Gabriela Imreh. 2002. “Practicing Perfection: Piano Performance as Expert Memory.” *Psychological Science* 13 (4): 342–49. https://doi.org/10.1111/j.0956-7976.2002.00462.x.

Chang, Sungkyun, and Kyogu Lee. 2014. “A Pairwise Approach to Simultaneous Onset/Offset Detection for Singing Voice Using Correntropy.” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, 629–33. doi:10.1109/ICASSP.2014.6853672.

Chu, Mingyuan, and Peter Hagoort. 2014. “Synchronization of Speech and Gesture: Evidence for Interaction in Action.” *Journal of Experimental Psychology: General* 143 (4): 1726–41. http://dx.doi.org/10.1037/a0036281.

Clayton, Martin. 2007a. “Time, Gesture and Attention in a ‘Khyāl’ Performance.” *Asian Music* 38 (2): 71–96. doi:10.1353/amu.2007.0032.

———. 2007b. “Observing Entrainment in Music Performance: Video-Based Observational Analysis of Indian Musicians’ Tanpura Playing and Beat Marking.” *Musicae Scientiae* 11 (1): 27–59. https://doi.org/10.1177/102986490701100102.

———. 2012. “What Is Entrainment? Definition and Applications in Musical Research.” *Empirical Musicology Review* 7 (1–2): 49–56. https://doi.org/10.18061/1811/52979.

———. 2015. “Aksak Patterns and Entrained Interaction in Transylvanian Village Music.” *Empirical Musicology Review* 10 (4): 292–301. http://dx.doi.org/10.18061/emr.v10i4.4886.

Clayton, Martin, Rebecca Sager, and Udo Will. 2005. “In Time with the Music: The Concept of Entrainment and Its Significance for Ethnomusicology.” *European Meetings in Ethnomusicology* 11: 1–82.

Colton, Raymond, and Edward G. Conture. 1990. “Problems and Pitfalls of Electroglottography.” *Journal of Voice* 4 (1): 10–24. https://doi.org/10.1016/S0892-1997(05)80077-3.

Covey-Crump, Roger. 1992. “Vocal Consort Style and Tuning.” In *Companion to Contemporary*

Musical Thought, 2, edited by John Paynter, 1020–50. London and New York: Routledge Reference.

Czeisler, Charles, James S. Allan, Steven H. Strogatz, Joseph M. Ronda, Ramiro Sánchez, David C. Ríos, Walter O. Freitag, Gary S. Richardson, and Richard E. Kronauer. 1986. “Bright Light Resets the Human Circadian Pacemaker Independent of the Timing of the Sleep-Wake Cycle.” *Science* 233 (4764): 667–71. https://doi.org/10.1126/science.3726555.

D’Amario, Sara, and Helena Daffern. 2017. “Using Electrolaryngography and Electroglottography to Assess the Singing Voice: A Systematic Review.” *Psychomusicology: Music, Mind and Brain* 27 (4): 229–43. http://dx.doi.org/10.1037/pmu0000184.

D’Amario, Sara, Helena Daffern, and Freya Bailes. 2018a. “A Longitudinal Study Investigating Synchronization in a Singing Quintet.” *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2018.07.015.

———. 2018b. “A New Method of Onset and Offset Detection in Ensemble Singing.” *Logopedics Phoniatrics Vocology*. https://doi.org/10.1080/14015439.2018.1452977.

———. 2018c. “Synchronization in Singing Duo Performances: The Roles of Visual Contact and Leadership Instruction.” *Frontiers in Psychology* 9 (July): 1208. https://doi.org/10.3389/fpsyg.2018.01208.

D’Amario, Sara, David M. Howard, Helena Daffern, and Nicola Pennill. 2018. “A Longitudinal Study of Intonation in an a Cappella Singing Quintet.” *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2018.07.015.

Daffern, Helena, and Jude Brereton. 2013. “Testing a New Protocol to Measure Tuning Response Behaviour in Solo Voice Ensemble Singing.” In *Proceedings of the Stockholm Music Acoustics Conference SMAC*, 231–37.

Dahl, Sofia, and Anders Friberg. 2007. “Visual Perception of Expressiveness in Musicians’ Body Movements.” *Music Perception: An Interdisciplinary Journal* 24 (5): 433–54. https://www.jstor.org/stable/10.1525/mp.2007.24.5.433.

Daugherty, James. 1999. “Spacing, Formation, Choral Preferences and Sound: Auditors

Choristers Perceptions of and Perceptions of Auditors and Choristers.” *Journal of Research in Music Education* 47 (3): 224–38. https://doi.org/10.2307/3345781.

Davidson, Jane W. 2001. “The Role of the Body in the Production and Perception of Solo Vocal Performance: A Case Study of Annie Lennox.” *Musicae Scientiae* 5 (2): 235–56. https://doi.org/10.1177/102986490100500206.

Davidson, Jane W., and Elaine King. 2004. “Strategies for Ensemble Practice.” In *Musical Excellence: Strategies and Techniques to Enhance Performance*, edited by Aaron Williamon, 105–22. Oxford: Oxford University Press.

Davies, P., Geoffrey A. Lindsey, Hilary Fuller, and Adrian J. Fourcin. 1986. “Variation in Glottal Open and Closed Phases for Speakers of English.” In *Proceedings of the Institute of Acoustics, 8*, 539–546.

de Almeida Bezerra, Adriana, Sabrina Cukier-Blaj, André Duprat, Zuleica Camargo, and Lídio Granato. 2009. “The Characterization of the Vibrato in Lyric and Sertanejo Singing Styles: Acoustic and Perceptual Auditory Aspects.” *Journal of Voice* 23 (6): 666–70. https://doi.org/10.1016/j.jvoice.2008.04.002.

de Cheveigné, Alain, and Hideki Kawahara. 2002. “YIN, a Fundamental Frequency Estimator for Speech and Music.” *The Journal of the Acoustical Society of America* 111 (4): 1917–30. https://doi.org/10.1121/1.1458024.

Dejonckere, Philippe, Minoru Hirano, and Johan Sundberg. 1995. *Vibrato*. San Diego: Singular Publishing Group.

Demos, Alexander P., Roger Chaffin, Kristen T. Begosh, Jennifer R. Daniels, and Kerry L. Marsh. 2012. “Rocking to the Beat: Effects of Music and Partner’s Movements on Spontaneous Interpersonal Coordination.” *Journal of Experimental Psychology: General* 141 (1): 49–53. http://dx.doi.org/10.1037/a0023843.

Devaney, Johanna. 2011. “An Empirical Study of the Influence of Musical Context on Intonation Practices in Solo Singers and SATB Ensembles. PhD Dissertation, McGill University.”

Devaney, Johanna, Michael Mandel, and Ichiro Fujinaga. 2012. “A Study of Intonation in

Three-Part Singing Using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT).” In *13th International Society for Music Information Retrieval Conference ISMIR 2012*, 511–16. http://ismir2012.ismir.net/event/papers/511-ismir-2012.pdf.

Devaney, Johanna, Michael I. Mandel, Daniel P. W. Ellis, and Ichiro Fujinaga. 2011. “Automatically Extracting Performance Data from Recordings of Trained Singers.” *Psychomusicology: Music, Mind and Brain* 21 (1–2): 108–36. http://dx.doi.org/10.1037/h0094008.

Dixon, Simon. 2001. “Learning to Detect Onsets of Acoustic Piano Tones.” In *Proceedings of the MOSART Workshop on Current Directions in Computer Music Research*, 147–51. Barcelona, Spain.

———. 2006. “Onset Detection Revisited.” *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, 1–6.

Doffman, Mark. 2008. “Feeling the Groove: Shared Time and Its Meanings for Three Jazz Trios.” PhD dissertation, Open University.

———. 2009. “Making It Groove! Entrainment, Participation and Discrepancy in the ‘Conversation’ of a Jazz Trio.” *Language and History* 52 (1): 130–47. https://doi.org/10.1179/175975309X452012.

———. 2013. “Groove: Temporality, Awareness and the Feeling of Entrainment in Jazz Performance.” In *Experience and Meaning in Music Performance*, edited by Martin Clayton, Byron Dueck, and Laura Leante, 62–85. Oxford: Oxford University Press.

Easley, Eleanor. 1932. “A Comparison of the Vibrato in Concert and Opera Singing.” In *University of Iowa Studies in the Psychology of Music, Vol. I: The Vibrato*, edited by Carl E. Seashore, 269–275. Iowa City: University of Iowa Press.

Eerola, Tuomas, Kelly Jakubowski, Nikki Moran, Peter E. Keller, and Martin Clayton. 2018. “Shared Periodic Performer Movements Coordinate Interactions in Duo Improvisations.” *Royal Society Open Science* 5: 171520. http://dx.doi.org/10.1098/rsos.171520.

Ellamil, Melissa, Joshua Berson, Jen Wong, Louis Buckley, and Daniel S. Margulies. 2016. “One in the Dance: Musical Correlates of Group Synchrony in a Real-World Club Environment.” *PLoS ONE* 11 (10): e0164783. https://doi.org/10.1371/journal.pone.0164783.

Fischinger, Timo, Klaus Frieler, and Jukka Louhivuori. 2015. "Influence of Virtual Room Acoustics on Choir Singing." *Psychomusicology: Music, Mind, and Brain* 25 (3): 208–218. http://dx.doi.org/10.1037/pmu0000117

Fourcin, Adrian, and Evelyn R. M. Abberton. 1971. “First Applications of a New Laryngograph.” *Medical and Biological Illustration* 21: 172–82.

Fredrickson, William E. 1994. “Band Musicians’ Performance and Eye Contact as Influenced by Loss of a Visual and/or Aural Stimulus.” *Journal of Research in Music Education* 42 (4): 306–17. https://doi.org/10.2307/3345738.

Friberg, Anders, Erwin Schoonderwaldt, and Patrik N. Juslin. 2007. “CUEX: An Algorithm for Automatic Extraction of Expressive Tone Parameters in Music Performance from Acoustic Signals.” *Acta Acustica United with Acustica* 93 (3): 411–20.

Fry, Dennis, and Lucie Manén. 1957. “Basis for the Acoustical Study of Singing.” *The Journal of the Acoustical Society of America* 29 (6): 690–92. https://doi.org/10.1121/1.1909011.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942.

Gersick, Connie J. G. 1988. “Time and Transition in Work Teams: Toward a New Model of Group Development.” *The Academy of Management Journal* 31 (1): 9–41. https://doi.org/10.5465/256496.

Ginsborg, Jane, Roger Chaffin, and George Nicholson. 2006. “Shared Performance Cues in Singing and Conducting: A Content Analysis of Talk during Practice.” *Psychology of Music* 34 (2): 167–94. https://doi.org/10.1177/0305735606061851.

Ginsborg, Jane, and Elaine King. 2007. “The Roles of Expertise and Partnership in Collaborative Rehearsal.” *Proceedings of the International Symposium on Performance Science*, 61–66. http://www.legacyweb.rcm.ac.uk/cache/fl0020172.pdf.

———. 2012. “Rehearsal Talk: Familiarity and Expertise in Singer-Pianist Duos.” *Musicae*

Scientiae 16 (2): 148–67. https://doi.org/10.1177/1029864911435733.

Glowinski, Donald, Leonardo Badino, Alessandro D'Ausilio, Antonio Camurri, and Luciano Fadiga. 2012. "Analysis of Leadership in a String Quartet." *Third International Workshop on Social Behaviour in Music at ACM ICMI 2012, Santa Monica, CA.*

Glowinski, Donald, Floriane Dardard, Giorgio Gnecco, Stefano Piana, and Antonio Camurri. 2015. "Expressive Non-Verbal Interaction in a String Quartet: An Analysis through Head Movements." *Journal on Multimodal User Interfaces* 9 (1): 55–68. https://doi.org/10.1007/s12193-014-0154-3.

Glowinski, Donald, Maurizio Mancini, Roddy Cowie, Antonio Camurri, Carlo Chiorri, and Cian Doherty. 2013. "The Movements Made by Performers in a Skilled Quartet: A Distinctive Pattern, and the Function That It Serves." *Frontiers in Psychology* 4 (November): 841. https://doi.org/10.3389/fpsyg.2013.00841.

Goebl, Werner. 2003. "The Role of Timing and Intensity in the Production and Perception of Melody in Expressive Piano Performance." PhD Dissertation, Karl-Franzens-Universität Graz, Austria. Retrieved from http://iwk.mdw.ac.at/goebl.

Goebl, Werner, Simon Dixon, and Emery Schubert. 2014. "Quantitative Methods: Motion Analysis, Audio Analysis, and Continuous Response Techniques." In *Expressiveness in Music Performance: Empirical Approaches across Styles and Cultures*, edited by Dorottya Fabian, Renee Timmers, and Emery Schubert, 221–39. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199659647.003.0013.

Goebl, Werner, and Caroline Palmer. 2009. "Synchronization of Timing and Motion Among Performing Musicians." *Music Perception* 26 (5): 427–38. http://dx.doi.org/10.1525/MP.2009.26.5.427.

Goebl, Werner, and Richard Parncutt. 2001. "Perception of Onset Asynchronies : Acoustic Piano versus Synthesized Complex versus Pure Tones." *Society* X (1993): 2001.

———. 2002. "The Influence of Relative Intensity on the Perception of Onset Asynchronies." *ICMPC7, July 2002*, no. July: 1–4. papers://eba95c07-21c6-4247-a2f7-57339f1ca15a/Paper/p11980.

———. 2003. “Asynchrony versus Intensity as Cues for Melody Perception in Chords and Real Music.” In *Proceedings of the 5th Triennial ESCOM Conference, Hanover, Germany*, edited by Reinhard Kopiez, Andreas Lehmann, and Irving Wolther, 376–380.

Goodman, Elaine. 2002. “Ensemble Performance.” In *Musical Performance: A Guide to Understanding*, edited by John Rink, 153–67. Cambridge: Cambridge University Press.

Hacki, Thamás. 1996. “Electroglottographic Quasi-Open Quotient and Amplitude in Crescendo Phonation.” *Journal of Voice* 10: 342–347. http://dx.doi.org/ 10.1016/S0892-1997(96)80025-7.

Hansen, Curtis. 1964. “Tuning the Choir.” *Music Educators Journal* 51 (2): 85–89. https://doi.org/10.2307/3390307.

Herbst, Christian, Harm K. Schutte, Daniel L. Bowling, and Jan G. Svec. 2017. “Comparing Chalk With Cheese—The EGG Contact Quotient Is Only a Limited Surrogate of the Closed Quotient.” *Journal of Voice* 31 (4): 401–9. http://dx.doi.org/10.1016/j.jvoice.2016.11.007.

Himberg, Tommi, and Marc Thompson. 2009. “Group Synchronization of Coordinated Movements in a Cross-Cultural Choir Workshop.” *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM)*, 175–80.

Hirano, Minoru, Seishi Hibi, and Tomoaki Sanada. 1989. “Falsetto, Head/Chest, and Speech Mode: An Acoustic Study with Three Tenors.” *Journal of Voice* 3 (2): 99–103. https://doi.org/10.1016/S0892-1997(89)80135-3.

Hirsh, Ira. 1959. “Auditory Perception of Temporal Order.” *The Journal of the Acoustical Society of America* 31 (6): 759–67. https://doi.org/10.1121/1.1907782.

Hofmann, Alex, Brian C. Wesolowski, and Werner Goebl. 2017. “The Tight-Interlocked Rhythm Section: Production and Perception of Synchronisation in Jazz Trio Performance.” *Journal of New Music Research* 46 (4). Routledge: 329–41. doi:10.1080/09298215.2017.1355394.

Hove, Michael J., Peter E. Keller, and Carol L. Krumhansl. 2007. “Sensorimotor Synchronization with Chords Containing Tone-Onset Asynchronies.” *Perception & Psychophysics* 69 (5):

699–708. https://doi.org/10.3758/BF03193772.

Howard, David M. 1995. "Variation of Electrolaryngographically Derived Closed Quotient for Trained and Untrained Adult Female Singers." *Journal of Voice* 9 (2): 163–72. http://dx.doi.org/10.1016/S0892-1997(05)80250-4.

———. 2007a. "Equal or Non-Equal Temperament in a Capella SATB Singing." *Logopedics, Phoniatrics, Vocology* 32 (2): 87–94. http://dx.doi.org/10.1080/14015430600865607.

———. 2007b. "Intonation Drift in a Capella Soprano, Alto, Tenor, Bass Quartet Singing with Key Modulation." *Journal of Voice* 21 (3): 300–315. http://dx.doi.org/10.1016/j.jvoice.2005.12.005.

———. 2007c. "Larynx Closed Quotient Variation in Quartet Singing." In *19 Th International Congress on Acoustics, Madrid, 2-7 September*.

———. 2009. "Electroglottogrphy/Electrolaryngography." In *The Larynx*, edited by Marvin Fried and Alfio Ferlito, 3rd ed., 227–43. San Diego: Plural Press.

Howard, David M., and Jamie A. S. Angus. 2017. *Acoustics and Psychoacoustics*. 5th ed. Oxford, UK: Focal Press.

Howard, David M., Jude Brereton, Graham Welch, Evangelos Himonides, Michael DeCosta, Jenevora Williams, and Andrew W. Howard. 2007. "Are Real-Time Displays of Benefit in the Singing Studio? An Exploratory Study." *Journal of Voice* 21 (1): 20–34. https://doi.org/10.1016/j.jvoice.2005.10.003.

Ifeachor, Emmanuel, and Barrie Jervis. 2002. *Digital Signal Processing: A Practical Approach, 2nd Edition*. USA: Pearson Education.

Issartel, Johann, Ludovic Marin, and Marielle Cadopi. 2007. "Unintended Interpersonal Co-Ordination: 'Can We March to the Beat of Our Own Drum?'" *Neuroscience Letters* 411 (3): 174–79. https://doi.org/10.1016/j.neulet.2006.09.086.

Jakubowski, Kelly, Tuomas Eerola, Paolo Alborno, Gualtiero Volpe, Antonio Camurri, and Martin Clayton. 2017. "Extracting Coarse Body Movements from Video in Music Performance: A Comparison of Automated Computer Vision Techniques with Motion Capture Data." *Frontiers in Digital Humanities* 4 (April): 1–10.

https://doi.org/10.3389/fdigh.2017.00009.

Janata, Petr, Stefan T. Tomic, and Jason M. Haberman. 2012. “Sensorimotor Coupling in Music and the Psychology of the Groove.” *Journal of Experimental Psychology: General* 141 (1): 54–75. doi:10.1037/a0024208.

Kawase, Satoshi. 2014. “Gazing Behavior and Coordination during Piano Duo Performance.” *Attention, Perception, & Psychophysics* 76 (2): 527–40. https://doi.org/10.3758/s13414-013-0568-0.

Keller, Peter E. 2008. “Joint Action in Music Performance.” In *Enacting Intersubjectivity: A Cognitive and Social Perspective to the Study of Interactions*, edited by Francesca Morganti, Antonella Carassa, and Giuseppe Riva, 205–21. Amsterdam: IOS Press.

———. 2013. “Musical Ensemble Performance: A Theoretical Framework and Empirical Findings on Interpersonal Coordination.” In *International Symposium on Performance Science, Vienna, 23-31 August 2013*, 271–85.

———. 2014. “Ensemble Performance: Interpersonal Alignment of Musical Expression.” In *Expressiveness in Music Performance: Empirical Approaches across Styles and Cultures*, edited by Dorottya Fabian, Renee Timmers, and Emery Schubert, 260–82. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199659647.003.0015.

Keller, Peter E., and Mirjam Appel. 2010. “Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles.” *Music Perception: An Interdisciplinary Journal* 28 (1): 27–46. https://doi.org/10.1525/mp.2010.28.1.27.

Keller, Peter E., Günther Knoblich, and Bruno H. Repp. 2007. “Pianists Duet Better When They Play with Themselves: On the Possible Role of Action Simulation in Synchronization.” *Consciousness and Cognition* 16 (1): 102–11. https://doi.org/10.1016/j.concog.2005.12.004.

Keller, Peter E., and Iring Koch. 2008. “Action Planning in Sequential Skills: Relations to Music Performance.” *The Quarterly Journal of Experimental Psychology* 61 (2): 275–91. doi:10.1080/17470210601160864.

Keller, Peter E., Giacomo Novembre, and Michael J. Hove. 2014. “Rhythm in Joint Action:

Psychological and Neurophysiological Mechanisms for Real-Time Interpersonal Coordination.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 369 (1658): 20130394. http://dx.doi.org/10.1098/rstb.2013.0394.

Kenny, Dianna, and Helen F. Mitchell. 2006. “Acoustic and Perceptual Appraisal of Vocal Gestures in the Female Classical Voice.” *Journal of Voice* 20 (1): 55–70. https://doi.org/10.1016/j.jvoice.2004.12.002.

———. 2007. “Vocal Quality in Female Classical Singers: The Role of Acoustics, Perception and Pedagogy.” In *Proceedings of the Third Conference on Interdisciplinary Musicology (CIM07) Tallinn, Estonia, 15-19 August 2007*, edited by Kaire Maimets-Volt, Richard Parncutt, Manuela Marin, and Jaan Ross.

King, Elaine. 2006. “The Roles of Student Musicians in Quartet Rehearsals.” *Psychology of Music* 34 (2): 262–82. https://doi.org/10.1177/0305735606061855.

Klapuri, Anssi. 1999. “Sound Onset Detection by Applying Psychoacoustic Knowledge.” *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing.* 6: 3089–92. doi:10.1109/ICASSP.1999.757494.

Kurosawa, Kaori, and Jane W. Davidson. 2005. “Nonverbal Behaviours in Popular Music Performance: A Case Study of The Corrs.” *Musicae Scientiae* 9 (1): 111–36. https://doi.org/10.1177/102986490500900104.

Leys, Christophe, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. “Detecting Outliers: Do Not Use Standard Deviation around the Mean, Use Absolute Deviation around the Median.” *Journal of Experimental Social Psychology* 49 (4): 764–66. http://dx.doi.org/10.1016/j.jesp.2013.03.013.

Lim, Mei Chern. 2014. “In Pursuit of Harmony: The Social and Organisational Factors in a Professional Vocal Ensemble.” *Psychology of Music* 42 (3): 307–24. https://doi.org/10.1177/0305735612469674.

Lindenberger, Ulman, Shu Chen Li, Walter Gruber, and Viktor Müller. 2009. “Brains Swinging in Concert: Cortical Phase Synchronization While Playing Guitar.” *BMC Neuroscience* 10: 22. https://doi.org/10.1186/1471-2202-10-22.

Loehr, Janeen, Edward W. Large, and Caroline Palmer. 2011. “Temporal Coordination and Adaptation to Rate Change in Music Performance.” *Journal of Experimental Psychology: Human Perception and Performance* 37 (4): 1292–1309. http://dx.doi.org/10.1037/a0023102.

Loehr, Janeen, and Caroline Palmer. 2009. “Subdividing the Beat: Auditory and Motor Contributions to Synchronization.” *Music Perception: An Interdisciplinary Journal* 26 (5): 415–25. https://doi.org/10.1525/mp.2009.26.5.415.

Lucas, Glaura, Martin Clayton, and Laura Leante. 2011. “Inter-Group Entrainment in Afro-Brazilian Congado Ritual.” *Empirical Musicology Review* 6 (2): 75–102. https://doi.org/10.18061/1811/51203.

Madison, Guy. 2006. “Experiencing Groove Induced by Music: Consistency and Phenomenology.” *Music Perception: An Interdisciplinary Journal* 24 (2): 201–8. https://www.jstor.org/stable/10.1525/mp.2006.24.2.201 Accessed:

Manfredi, Claudia, Davide Barbagallo, Giovanna Baracca, Silvia Orlandi, Andrea Bandini, and Philippe H. Dejonckere. 2015. “Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers.” *Journal of Voice* 29 (4): 517.e1-517.e9. http://dx.doi.org/10.1016/j.jvoice.2014.09.014.

Mardia, Kanti, and Peter Jupp. 2000. *Directional Statistics*. Chichester, UK: John Wiley & Sons.

Markel, John. 1972. “The SIFT Algorithm for Fundamental Frequency Estimation.” *IEEE Transactions on Audio and Electroacoustics* 20 (5): 367–77. doi:10.1109/TAU.1972.1162410.

Masri, Paul. 1996. “Computer Modeling of Sound for Transformation and Synthesis of Musical Signal.” PhD dissertation, University of Bristol.

Metfessel, Milton. 1929. “The Strobophotograph: A Device for Measuring Pitch.” *Journal of General Psychology* 2 (1): 135–39. https://doi.org/10.1080/00221309.1929.9918047.

———. 1932. “The Vibrato in Artistic Voices.” In *University of Iowa Studies in the Psychology of Music, Vol. I: The Vibrato*, edited by Carl E. Seashore, 14–117. Iowa City: University of Iowa Press.

Meyns, Pieter, Sjoerd M. Bruijn, and Jacques Duysens. 2013. “The How and Why of Arm Swing during Human Walking.” *Gait & Posture* 38 (4): 555–62. https://doi.org/10.1016/j.gaitpost.2013.02.006.

Miles, Walter. 1956. *Carl Emil Seashore, 1866-1949*. New York: Columbia University Press.

Miller, Dayton. 1916. “The Henrici Harmonic Analyzer and Devices for Extending and Facilitating Its Use.” *Journal of the Franklin Institute* 182 (3): 285–322. https://doi.org/10.1016/S0016-0032(16)90888-3.

Miller, Ray. 1936. “The Pitch of the Attack in Singing.” In *University of Iowa Studies in the Psychology of Music, Vol. IV: Objective Analysis of Muscial Performance*, edited by Carl E. Seashore, 158–171. Iowa City: University of Iowa Press.

MIREX. 2016. “Onset Detection F-Measure per Class.” http://nema.lis.illinois.edu/nema_out/mirex2016/results/aod/ resultsperclass.html.

Morris, Richard, Sten Ternström, Jeannette Lovetri, and Dianne Berkun. 2012. “Long-Term Average Spectra from a Youth Choir Singing in Three Vocal Registers and Two Dynamic Levels.” *Journal of Voice* 26 (1): 30–36. https://doi.org/10.1016/j.jvoice.2010.07.003%0A.

Mossbridge, Julia, Matthew B. Fitzgerald, Erin S. O’Connor, and Beverly A. Wright. 2006. “Perceptual-Learning Evidence for Separate Processing of Asynchrony and Order Tasks.” *Journal of Neuroscience* 26 (49): 12708–16. https://doi.org/10.1523/JNEUROSCI.2254-06.2006.

Müller, Viktor, and Ulman Lindenberger. 2011. “Cardiac and Respiratory Patterns Synchronize between Persons during Choir Singing.” *PLoS ONE* 6 (9). https://doi.org/10.1371/journal.pone.0024893.

Néda, Zoltán, Ercsey Ravasz, Yves Brechet, Tamás Vicsek, and Albert Laszlo Barabási. 2000. “Self-Organizing Processes: The Sound of Many Hands Clapping.” *Nature* 403 (6772): 849–850. https://doi.org/10.1038/35002660.

Nusseck, Manfred, and Marcelo Wanderley. 2009. “Music and Motion—How Music-Related Ancillary Body Movements Contribute to the Experience of Music.” *Music Perception: An*

Interdisciplinary Journal 26 (4): 335–53. doi:10.1111/j.1835-9310.1982.tb01239.x.

Oullier, Olivier, Gonzalo C. de Guzman, Kelly J. Jantzen, Julien Lagarde, and Scott J. A. Kelso. 2008. “Social Coordination Dynamics: Measuring Human Bonding.” *Social Neuroscience* 3 (2): 178–92. https://doi.org/10.1080/17470910701563392.

Pabst, Friedemann, and Johan Sundberg. 1993. “Tracking Multi-Channel Electroglottograph Measurement of Larynx Height in Singers.” *Scandinavian Journal of Logopedics and Phoniatrics* 18 (4): 143–52. https://doi.org/10.3109/14015439309101360.

Page-Shipp, Roy, Down Joseph, and Caroline van Niekerk. 2018. “Conductorless Singing Group: A Particular Kind of Self-Managed Team?” *Team Performance Management: An International Journal* 24 (5/6): 331–46. https://doi.org/10.1108/TPM-09-2016-0041.

Palmer, Caroline, Frances Spidle, Erik Koopmans, and Peter Schubert. 2013. “Temporal Coordination in Vocal Duet Performances of Musical Rounds.” In *Proceedings of the Stockholm Music Acoustics Conference 2013, SMAC 2013, Stockholm, Sweden Temporal*, 678–82.

Pastore, Richard, Laura B. Harris, and Jody K. Kaplan. 1982. “Temporal Order Identification: Some Parameter Dependencies.” *The Journal of the Acoustical Society of America* 71 (2): 430–36. doi:10.1121/1.387446.

Phillips-Silver, Jessica, and Peter E. Keller. 2012. “Searching for Roots of Entrainment and Joint Action in Early Musical Interactions.” *Frontiers in Human Neuroscience* 6 (February): 26. https://doi.org/10.3389/fnhum.2012.00026.

Pikovsky, Arkady, Michael Rosenblum, and Jurgen Kurths. 2001. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511755743.002.

Prame, Eric. 1994. “Measurements of the Vibrato Rate of Ten Singers.” *The Journal of the Acoustical Society of America* 96 (4): 1979–84. https://doi.org/10.1121/1.410141.

Ragert, Marie, Tim Schroeder, and Peter E. Keller. 2013. “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* 4 (June): 368.

https://doi.org/10.3389/fpsyg.2013.00368.

Rasch, Rudolf. 1979. “Synchronization in Performed Ensemble Music.” *Acta Acustica United with Acustica* 43 (2): 121–31.

———. 1988. “Timing and Synchronisation in Ensemble Performance.” In *Generative Processes in Music*, edited by John Sloboda, 70–90. Oxford: Oxford University Press.

Repp, Bruno H. 2001. “Processes Underlying Adaptation to Tempo Changes in Sensorimotor Synchronization.” *Human Movement Science* 20 (3): 277–312. https://doi.org/10.1016/S0167-9457(01)00049-5.

———. 2005. “Sensorimotor Synchronization: A Review of the Tapping Literature.” *Psychonomic Bulletin & Review* 12 (6): 969–92. https://doi.org/10.3758/BF03206433.

Repp, Bruno H., and Peter E. Keller. 2004. “Adaptation to Tempo Changes in Sensorimotor Synchronization: Effects of Intention, Attention, and Awareness.” *Quarterly Journal of Experimental Psychology* 57A (3): 499–521. https://doi.org/10.1080/02724980343000369.

Repp, Bruno H., and Yi-Huang Su. 2013. “Sensorimotor Synchronization: A Review of Recent Research (2006-2012).” *Psychonomic Bulletin & Review* 20 (3): 403–52. https://doi.org/10.3758/s13423-012-0371-2.

RStudio. 2015. “RStudio: Integrated Development for R.” Boston: RStudio Inc. http://www.rstudio.com/.

Sacerdote, Gino. 1957. “Researches on the Singing Voice.” *Acta Acustica United with Acustica* 7 (2): 61–68.

Sänger, Johanna, Viktor Müller, and Ulman Lindenberger. 2012. “Intra- and Interbrain Synchronization and Network Properties When Playing Guitar in Duets.” *Frontiers in Human Neuroscience* 6 (November): 1–19. https://doi.org/10.3389/fnhum.2012.00312.

———. 2013. “Directionality in Hyperbrain Networks Discriminates between Leaders and Followers in Guitar Duets.” *Frontiers in Human Neuroscience* 7 (June): 1–14. https://doi.org/10.3389/fnhum.2013.00234.

Scherer, Ronald, David G. Druker, and Ingo R. Titze. 1988. “Electroglottography and Direct

Measurement of Vocal Fold Contact Area." In *Vocal Fold Physiology: Voice Production, Mechanisms and Functions*, edited by Osamu Fujimura, 279–90. New York: Raven Press.

Schoen, Max. 1922. "An Experimental Study of the Pitch Factor in Artistic Singing." *Psychological Monographs* 31: 230–259.

Seashore, Carl E. 1914. "The Tonoscope." *Psychological Monographs* 16: 1–12.

———. 1929. *The Measurement of Pitch Intonation with the Tonoscope in Singing and Playing*. Iowa City: University of Iowa Press.

———. 1936. *Objective Analysis of Musical Performance*. Iowa City: University of Iowa Press.

———. 1938. *Psychology of Music*. New York: McGraw- Hill.

———. 1947. *In Search of Beauty in Music: A Scientific Approach to Musical Esthetics*. New York: Ronald Press.

Silva, Pedro, Dante Chung, Tiago Cardoso, Carvalho Thiago, Keith Davids, Araújo Duarte, and Júlio Garganta. 2016. "Practice Effects on Intra-Team Synergies in Football Teams." *Human Movement Science* 46: 39–51. https://doi.org/10.1016/j.humov.2015.11.017.

Strogatz, Steven, and Ian Stewart. 1993. "Coupled Oscillators and Biological Synchronization." *Scientific American* 269 (6): 102–9.

Sundberg, Johan. 1987. *The Science of the Singing Voice*. Illinois: Northern Illinois University Press.

———. 2001. "Level and Center Frequency of the Singer's Formant." *Journal of Voice* 15 (2): 176–86. https://doi.org/10.1016/S0892-1997(01)00019-4.

Sundberg, Johan, and Julia Bauer-Huppmann. 2007. "When Does a Sung Tone Start?" *Journal of Voice* 21 (3): 285–93. doi:10.1016/j.jvoice.2006.01.003.

Sundberg, Johan, Eric Prame, and Jenny Iwarsson. 1995. "Replicability and Accuracy of Pitch Patterns in Professional Singers." *Dept. for Speech, Music and Hearing. Quartely Progress and Status Report* 36: 51–62.

Ternström, Sten. 2003. "Choir Acoustics–an Overview of Scientific Research Published to Date." *International Journal of Research in Choral Singing* 43: 1–8.

http://www.speech.kth.se/prod/publications/files/qpsr/2002/2002_43_1_001-008.pdf.

Ternström, Sten, Sara D'Amario, and Andreas Selamtzis. 2018. "Effects of the Lung Volume on the Electroglottographic Waveform in Trained Female Singers." *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2018.09.006.

Ternström, Sten, and Jan Nordmark. 1996. "Intonation Preferences for Major Thirds with Non-Beating Ensemble Sounds." In *Nordic Acoustical Meeting, Helsinki, 12-14 June 1996*, 359–65.

Thalén, Margareta, and Johan Sundberg. 2001. "Describing Different Styles of Singing: A Comparison of a Female Singer's Voice Source in 'Classical', 'Pop', 'Jazz' and 'Blues'." *Logopedics, Phoniatrics, Vocology* 26 (2): 82–93. http://dx.doi.org/10.1080/140154301753207458.

Tiffin, Joseph. 1932. "Phonophotograph Apparatus." In *University of Iowa Studies in the Psychology of Music: Vol. 1. The Vibrato*, edited by Carl E. Seashore, 118–133. Iowa City: University of Iowa Press.

Timmers, Renee, Satoshi Endo, Adrian Bradbury, and Alan Wing. 2014. "Synchronization and Leadership in String Quartet Performance: A Case Study of Auditory and Visual Cues." *Frontiers in Psychology* 5 (June): 645. https://doi.org/10.3389/fpsyg.2014.00645.

Timmers, Renee, Satoshi Endo, and Alan Wing. 2013. "Temporal Coordination in String Quartet Performance." In *International Symposium on Performance Science*, 569–74. http://mediatum.ub.tum.de/doc/1207051/1207051.pdf.

Toh, Chee Chuan, Bingjun Zhang, and Ye Wang. 2008. "Multiple-Feature Fusion Based Onset Detection for Solo Singing Voice." In *9th International Conference on Music Information Retrieval*, edited by Juan P. Bello, Elaine Chew, and Douglas Turnbull, 515–20. Philadelphia: Drexel University.

Vickhoff, Björn, Helge Malmgren, Rickard Åström, Gunnar F. Nyberg, Seth-Reino Ekström, Mathias Engwall, Johan Snygg, Michael Nilsson, and Rebecka Jörnsten. 2013. "Music Structure Determines Heart Rate Variability of Singers." *Frontiers in Psychology* 4 (July): 334. https://doi.org/10.3389/fpsyg.2013.00334.

Wallach, Hans, Edwin B. Newman, and Mark R. Rosenzweig. 1949. "The Precedence Effect in Sound Localization." *The American Journal of Psychology* 62 (3): 315–36. http://www.jstor.org/stable/10.2307/1418275.

Wanderley, Marcelo, Bradley W. Vines, Neil Middleton, Cory McKay, and Wesley Hatch. 2005. "The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field." *Journal of New Music Research* 34 (1): 97–113. https://doi.org/10.1080/09298210500124208.

Wannier, Thierry, Carla Bastiaanse, Gery Colombo, and Volker Dietz. 2001. "Arm to Leg Coordination in Humans during Walking, Creeping and Swimming Activities." *Experimental Brain Research* 141 (3): 375–79. https://doi.org/10.1007/s002210100875.

Welch, Graham, David M. Howard, and Christine Rush. 1989. "Real-Time Visual Feedback in the Development of Vocal Pitch Accuracy in Singing." *Psychology of Music* 17 (2): 46–157. https://doi.org/10.1177/0305735689172005.

Welch, Graham, and Peta White. 1993. "The Developing Voice: Education and Vocal Efficiency. A Physical Perspective." *Bulletin of the Council for Research in Music Education* 119: 146–56. https://www.jstor.org/stable/40318623.

Widholm, Gregor. 2012. "Music Acoustics. A Recent Field of Research with Roots in Ancient Greece." In *5th Congress of Alps-Adria Acoustics Association, 12-14 September 2012, Petrčane, Croatia*, 1–9.

Williamon, Aaron, and Jane W. Davidson. 2002. "Exploring Co-Performer Communication." *Musicae Scientiae* 6 (1): 53–72. https://doi.org/10.1177/102986490200600103.

Wing, Alan, Satoshi Endo, Adrian Bradbury, and Dirk Vorberg. 2014. "Optimal Feedback Correction in String Quartet Synchronization." *Journal of the Royal Society Interface* 11: 20131125. http://dx.doi.org/10.1098/rsif.2013.1125.

Wing, Alan, Satoshi Endo, Tim Yates, and Adrian Bradbury. 2014. "Perception of String Quartet Synchronisation." *Frontiers in Psychology* 5 (October): 1115. https://doi.org/10.3389/fpsyg.2014.01115.

Yasuma, Fumihiko, and Jun-Ichiro Hayano. 2004. "Respiratory Sinus Arrhythmia - Why Does the Heartbeat Synchronize with Respiratory Rhythm?” *Chest* 125 (2): 683–690. https://doi.org/10.1378/chest.125.2.683.

Zamm, Anna, Peter Q. Pfordresher, and Caroline Palmer. 2015. “Temporal Coordination in Joint Music Performance: Effects of Endogenous Rhythms and Auditory Feedback.” *Experimental Brain Research* 233: 607–15. https://doi.org/10.1007/s00221-014-4140-5.

Zera, Jan, and David M. Green. 1993. “Detecting Temporal Onset and Offset Asynchrony in Multicomponent Complexes.” *The Journal of the Acoustical Society of America* 93 (2): 1038–52. https://doi.org/10.1121/1.405552.

Zhou, Ruohua, and J.D. Reiss. 2007. “Music Onset Detection Combining Energy-Based and Pitch-Based Approaches.” *Proc. MIREX Audio Onset Detection Contest*. doi:10.1.1.375.4365.