UNIVERSITY OF GHENT AND UNIVERSITY OF TURIN

CO-TUTELLE DOCTORATE: ART SCIENCE (GHENT) AND NEUROSCIENCE (TURIN)

DOCTORAL SCHOOL:

ARTS, HUMANITIES AND LAW (GHENT)

LIFE AND HEALTH SCIENCES (TURIN), CYCLE: XXXI

Brain and behavioural aspects of embodied musical interaction in dyads

CANDIDATE: ALESSANDRO DELL'ANNA

SUPERVISORS: PROF. MARC LEMAN (GHENT) AND PROF. ANNA BERTI (TORINO)

ACADEMIC YEARS: 2016-2019

TABLE OF CONTENTS

ACKNOWLEDGEMENT

Five years ago, after almost ten years involved as a special needs and/or philosophy teacher in the Italian high school and after recording a jazz album, I felt that it was the time to go back to research. Before those ten years, I had got a PhD in Philosophy of (Cognitive) Science (Genova University), investigating the roots of a new emerging cognitive paradigm, the so-called embodied mind framework (specifically in visual perception). Hence, it followed as a natural consequence to undertake this new exploration trying to combine my philosophical sensorimotor background with some new (to me) fields like musicology and neuroscience. One year later, Professors Anna Berti and Marc Leman turned out to be adventurous enough to plunge into this risky enterprise and I'm grateful to them, not only for sharing their outstanding scientific profile, but also because of their theoretical open-mindedness. In particular, I found the IPEM group in Gent a great crossroad of ideas, where people with radically different backgrounds meet and grow together, quite in the best spirit of the European Union. Also the collaboration of the SAMBA group's members in Torino has been precious. Moreover, Gent, as Torino, is such a lovely town that being there for research was always a pleasure. Of course, such an interdisciplinary work has its shortcomings and I'm sure that the experts in each field will spot them easily. Nevertheless, the same experts might benefit, I think, by a view from elsewhere, particularly those of them who are more theoretically-minded. It has been a dense period, further intensified by the birth of my second daughter. Although many people have filled my life during these years (relatives, friends, musicians, researchers), I can't but dedicate this effort to my family, Rossana, Naima and Sibilla (for *their* efforts and love).

English Summary

In the first chapter of this thesis, an Introduction to the driving concepts around the issue of musical interaction in dyads is presented. The concepts of "joint action", "musical interaction", "embodied cognition", "embodied music cognition", "expressive interaction" are discussed in order to pave the way to the empirical studies that follow. Finally, a paragraph is devoted to the neural underpinnings of musical interaction and an overview of the thesis through the open questions in the literature and the relative research questions this work deals with is offered. A central concept emerged from this theoretical introduction, that is, music (and, before it, musicality) as an embodied language, an intrinsically social device for coordination.

In the second chapter, I present a study in which, by means of alternate joint tapping task, we discovered that also pairs of non musicians are able to adapt their timing to their partner's, a competence that was held to be a musicians' exclusive so far. We did it by computing their asynchronies with respect to the metronome and cross-correlating them, finding that subjects tended to imitate their partner's asynchronies, rather than correcting for them, whether they occurred sooner or later than the metronome beats. Moreover, taking advantage of a well-known illusion of ownership (the rubber hand illusion), we found that mutual adaptive timing was established not only when the subject faced the partner (allocentric), but even in a condition in which the partner's hand is embodied (egocentric). We interpreted such a finding as corroborating the idea of a universal nature of musicality as an embodied language, due to the intrinsic characteristics of the joint proto-musical system. However, the physiological results collected thanks to single-pulse TMS enhance the conclusion that the "allocentric" condition was judged as the social one, resulting in higher cortico-spinal excitability than both the solo and the egocentric conditions. On the contrary, the embodiment of the partner led the subject to cope with the latter condition as if no partner was present at all, resulting in a cortico-spinal activation comparable to the solo (and the baseline) condition.

In the third chapter I explore the possibility that a musical interaction might remap the peri-personal space of two musicians according to the nature of such interaction, be it cooperative or uncooperative. It turned out that, after a jazz performance with an uncooperative partner, who played the wrong harmonic sequence of a jazz tune, a musician's peri-personal space shrank, as if a musical joint action became impossible. We interpreted this result as evidence that, insofar as music and musicality are intrinsically social, a musical interaction has a measurable impact on the perception of the space between two (or more) subjects. Consistently with an "extended mind" hypothesis, as tools can be incorporated into their users' body-schema and as body parts of another person can be embodied, we speculate that the peri-personal spaces of two (or more) interacting subjects may merge, as portrayed by the concept of "mutual incorporation". The audio-tactile integration task we exploited to measure peri-personal space allowed us to compare our sample of jazz musicians with a sample of non-musicians in the experimental baseline condition. As expected, the former group exhibited higher multisensory integration competences, arguably due to the musicians' sensorimotor training with their instrument and (to a lesser extent) singing, that brings about well-known cortical-subcortical reorganizations.

In the fourth chapter I tested a sample of wind musicians, under the hypothesis that their cortico-spinal excitability in FDI M1, triggerd by single-pulse TMS, could reveal their expertise in integrating (tactile) holding of a trumpet with the sound of a trumpet. Given the small sample, our results have to be considered as temporary, showing a trend toward a lower cortico-spinal excitability in musicians, compared to non-musicians, while listening to white noise, compared to trumpet tones, independently of the tool held. The lower sensitivity to white noise exhibited by musicians may be taken as a marker of their sensorimotor expertise, but further testing is needed to draw stronger conclusions about both populations' multisensory competences.

In the fifth chapter a hocket singing paradigm was devised, in which pairs of musicians' expressive quality was tested exploiting a Bayesian analysis method. Such a method allowed us to explore timing without assuming stationarity in the performance, under the hypothesis that predictive models carry out a continuous updating of timing-error minimisation, comparing the actual inter-onset intervals between the two singers' tones with the predicted ones. Interestingly, the movement condition showed a different effect according to the expertise, in that, while all the pairs tended to be more correct in the micro-timing (what we called, "fluctuation" errors), lower quality pairs disrupted significantly more often the performance in the movement, compared to the non-movement, condition (they made more "collapse" errors). From the subjective viewpoint, the singers reported a feeling of SHARED, rather than WE, agency, meaning that they felt to control only part of the joint action, rather than feeling as a single entity with their partner while accomplishing that action. However, it turned out that this subjective parameter was correlated with the objective parameter devised by our Bayesian analysis, i.e. duration errors (fluctuation, narration and collapse errors), showing its plausibility as a (subjective) marker of the expressive interaction. Even more highly correlated than joint agency with the timing markers of the interaction quality were the self-annotations of that very quality made by the musicians.

In the last chapter I discuss the results of our studies, answering the open questions posed in the Introduction and putting forward some questions for future research. A model of music (and musicality) as embodied language is sketched, aiming at integrating predictive coding and embodied frameworks.

Dutch Summary

In het eerste hoofdstuk van de thesis worden de basisconcepten inzake muziekinteractie in muzikantenparen toegelicht. De concepten van "gezamenlijke actie", "muziekinteractie", "lijfelijke cognitie", en "lijfelijke muziekcognitie" worden verduidelijkt in het licht van de empirische studies die volgen. Er wordt een paragraaf gewijd aan neuronale grondslagen van muziekinteractie en daarna wordt een overzicht gegeven van de thesis, dit ten aanzien van een aantal open vragen in de literatuur en van een aantal onderzoeksvragen die in de thesis aan bod komen. Een centraal concept dat naar voor komt in deze theoretische introductie is dat van muziek (en muzikaliteit) als lijfelijke taal en als intrinsiek sociaal instrument voor coördinatie.

In het tweede hoofdstuk presenteer ik een studie waarin, door middel van een alternerende tik-taak, gevonden werd dat paren van niet-muzikanten in staat zijn om hun tijdscoördinatie te adapteren aan hun partner. Dat is een competentie waarvan tot op heden gedacht werd dat enkel muzikanten daartoe in staat waren. We toonden dit aan door hun a-synchronisatie te berekenen met betrekking tot de metronoom en deze vervolgens te kruis-correleren. Zo vonden we dat subjecten hun partners a-synchronie neigen te imiteren, eerder dan hen te corrigeren wanneer ze vroeger of later waren dan de tikken van de metronoom. Door te werken met een gekende illusie van eigenaarschap (de rubberhand illusie) vonden we dat wederzijdse adaptieve tijdscoördinatie tot stand gebracht werd, niet enkel wanneer het subject tegenover de partner zat (allocentrisch), maar ook in een conditie waarin de partners hand lijfelijk gesimuleerd werd (egocentrisch). We interpreteren zo'n bevinding als de bevestiging van de idee dat muzikaliteit als lijfelijke taal universeel is, dankzij de intrinsieke eigenschappen van het proto-muzikaal systeem van elkeen. De fysiologische resultaten die we met enkel-puls TMS verzamelden versterken de conclusie dat de allocentrische conditie de sociale conditie was, wat resulteerde in een hogere cortico-spinale prikkeling dan bij solo en egocentrische condities. De lijfelijkheid van de partner leidde ertoe dat het subject met de laatstgenoemde conditie omging alsof geen partner aanwezig was, waardoor de cortico-spinale activatie vergelijkbaar was met de solo (en dus: "baseline") conditie.

In het derde hoofdstuk exploreer ik de mogelijkheid dat muzikale interactie aanleiding kan geven tot een herschikking van de peri-persoonlijke ruimte van twee muzikanten en dat volgens een interactie die hetzij coöperatief, hetzij niet-coöperatief is. Na een jazz uitvoering met een niet-coöperatieve partner, die de verkeerde harmonische akkoorden van een jazz melodie speelde, kromp de peri-persoonlijke ruimte van de muzikant, alsof de muzikale gezamenlijke actie onmogelijk werd. We interpreteerden dit resultaat door te stellen dat muziek en muzikaliteit intrinsiek sociaal zijn, waardoor muziekinteractie een meetbare impact heeft op de perceptie van de ruimte tussen twee (of meer) subjecten. In overeenkomst met een "extended mind" hypothese kunnen werktuigen geïncorporeerd worden in lijfelijk schema van de gebruiker. Deze werktuigen kunnen vervolgens als lichaamseigen instrumenten worden ervaren. Aangezien werktuigen op een lichamelijke manier kunnen toegeëigend worden veronderstellen we dat de peri-persoonlijke ruimtes van twee (of meer) interagerende subjecten inderdaad kunnen vermengen, zoals opgevat in het concept van wederzijdse belichaming. We hanteerden een audio-tactiele integratietaak om de peri-persoonlijke ruimte te meten en dat liet ons toe om een sample van jazz muzikanten te vergelijken met een sample van niet-muzikanten in de experimentele baseline-conditie. Zoals verwacht vertoonde de eerste groep een hogere multi-sensorische integratiecompetentie, wat verklaarbaar is door de eerdere sensorimotorische training van muzikanten met hun instrument en (in mindere mate) zingen, wat een gekende corticale-subcorticale reorganisatie impliceert.

In het vierde hoofdstuk testte ik een sample van blaasmuzikanten (trompettisten), onder de hypothese dat cortico-spinale prikkeling in FDI gerelateerd aan trompet-tactiliteit opgewekt wordt bij luisteren naar de klank van een trompet. Gegeven het kleine aantal subjecten zijn onze resultaten voorlopig, maar ze tonen een trend naar lagere cortico-spinale prikkeling in muzikanten, vergeleken bij niet-muzikanten, bij het luisteren naar witte ruis, vergeleken met trompet tonen, onafhankelijk van hoe het instrument wordt vastgehouden. De lagere sensitiviteit voor witte ruis bij muzikanten kan opgevat worden als een marker voor hun sensorimotorische expertise, maar verdere tests zijn nodig om sterkere conclusies over multi-sensorische competenties bij beide populaties te kunnen trekken.

In het vijfde hoofdstuk werd een zangparadigma uitwerkt, gebaseerd op hoketus zingen bij paren van muzikanten, en daarbij werd een Bayesiaanse analysemethode ontwikkeld. Zo'n methode liet ons toe om timing te exploreren zonder dat daarbij stationariteit hoeft verondersteld te worden. De hypothese is dat op basis van predictieve modellen die subjecten maken er een continue update gerealiseerd wordt van de timing-fout met de bedoeling om deze fout te minimaliseren. Daarbij worden de gezongen inter-aanzet intervallen tussen de twee zangers hun tonen vergeleken met de voorspelde inter-aanzet intervallen. De bewegingsconditie toonde een verschillend effect afhankelijk van de expertise. Terwijl de paren tendeerden naar een meer correcte micro-timing (fluctuatiefouten) werden paren die eerder laag scoren qua timing accuratesse significant meer beïnvloed door beweging, vergeleken met de conditie waarin ze stil stonden. Ze maakten daarbij meer "collapse"-fouten. Vanuit een subjectief standpunt gezien, rapporteerden de zangers een gevoel van "verbondenheids-", eerder dan een "wij-"gevoel van controle ("agency"). Dat betekent ze het gevoel hadden dat ze slechts een deel van de gezamenlijke actie onder controle hadden, eerder dan te voelen dat ze opgingen in een singuliere entiteit samen met de partner waarmee ze de actie tot een goed einde brengen. Echter, we vonden ook dat deze subjectieve parameter gecorreleerd was met de objectieve parameter geëxtraheerd uit onze Bayesiaanse analyse, dat is, duurtijdfouten, waarbij we een onderscheid maken tussen fluctuatie-, narratien collaps-fouten. Deze subjectieve parameter zou kunnen beschouwd worden als een (subjectieve) marker van de expressieve interactie. Daarbij bleek dat de zelf-annotaties van de kwaliteit van de uitvoering, door de muzikanten, zelfs hoger correleerden met de timing markers van de interactie dan het gezamenlijk gevoel van controle.

In het laatste hoofdstuk bespreek ik de resultaten van de studies. Daarbij beantwoord ik de open vragen die in de Inleiding ter sprake kwamen en stel ik een paar vragen voor toekomstig onderzoek. En model van muziek (en muzikaliteit) als lijfelijk taal wordt geschetst. Daarbij streef ik naar een integratie van de benaderingen inzake predictief coderen en lijfelijkheid.

Italian Summary

Nel primo capitolo di questa tesi presento un'Introduzione ai concetti portanti nella trattazione dell'interazione musicale in coppie. Allo scopo di preparare il campo agli studi empirici che seguono, discuto i concetti di "azione congiunta", "interazione musicale", "cognizione incorporata", "cognizione musicale incorporata" e "interazione espressiva". Infine, dedico un paragrafo alle basi neuronali dell'interazione musicale e offro una panoramica della tesi mediante le domande aperte in letteratura e le relative domande di ricerca affrontate. Da questa introduzione teorica emerge un concetto centrale, quello di musica (e, prima ancora, musicalità) come linguaggio incorporato, un dispositivo di coordinazione intrinsecamente sociale.

Nel secondo capitolo presento uno studio nel quale, per mezzo di un compito di tamburellamento alternato, abbiamo scoperto che anche coppie di non musicisti sono in grado di adattare il proprio timing a quello del partner, una competenza ritenuta finora prerogativa dei musicisti. L'abbiamo fatto calcolando le loro asincronie rispetto al metronomo e correlandole, trovando che i soggetti tendono a imitare le asincronie del proprio partner, piuttosto che correggerle, sia che esse siano in anticipo sia che siano in ritardo rispetto ai battiti del metronomo. Inoltre, approfittando di una ben nota illusione di "ownership" (l'illusione della mano di gomma), abbiamo scoperto che il mutuo adattamento del timing avveniva non solo quando il soggetto si trovava di fronte al partner (allocentrico), ma anche in una condizione in cui il braccio del partner veniva incorporato (egocentrica). Abbiamo interpretato questa scoperta come supporto all'idea di una natura universale della musicalità in quanto linguaggio incorporato, dovuta alle caratteristiche intrinseche del sistema proto-musicale congiunto. Comunque, i risultati fisiologici ottenuti grazie alla TMS a impulso singolo rafforzano la conclusione per cui la condizione allocentrica era ritenuta dai soggetti la condizione sociale, risolvendosi in un'eccitabilità cortico-spinale maggiore sia della condizione individuale che di quella egocentrica. Al contrario, l'incorporazione del partner portava il soggetto a guardare quest'ultima come se nessun partner fosse presente, traducendosi in un'attivazione cortico-spinale comparabile alla condizione individuale (e alla baseline).

Nel terzo capitolo esploro la possibilità che un'interazione musicale possa rimappare lo spazio peripersonale di due musicisti, a seconda della natura di tale interazione, sia essa cooperativa o non cooperativa. E' emerso che dopo una performance jazz con un partner non cooperativo, il quale suonava la sequenza armonica scorretta di un brano jazz, lo spazio peripersonale del musicista si restringesse, come se l'azione congiunta diventasse impossibile. Abbiamo interpretato questo risultato come prova che, nella misura in cui musica e musicalità sono intrinsecamente sociali, un'interazione musicale ha un impatto misurabile nella percezione dello spazio tra due (o più) soggetti. Coerentemente con un'ipotesi di "mente estesa", così come un attrezzo può essere incorporato nello schema corporeo di chi lo usa e così come parti del corpo di un'altra persona possono essere incorporate nel proprio, abbiamo immaginato che gli spazi peripersonali di due (o più) soggetti che interagiscono possano fondersi, come indicato dal concetto di "mutua incorporazione". Il compito d'integrazione uditivo-tattile che abbiamo impiegato per misurare lo spazio peripersonale ci ha consentito di confrontare il nostro campione di musicisti jazz con un campione di non musicisti nella condizione sperimentale di baseline. Come atteso, il primo gruppo ha mostrato competenze di integrazione

multisensoriale migliori, dovute probabilmente alla pratica sensomotoria dei musicisti con il loro strumento e (in misura minore) col canto, che provoca ben note riorganizzazioni cortico-sotto-corticali.

Nel quarto capitolo ho esaminato un campione di musicisti a fiato, nell'ipotesi che la loro eccitabilità cortico-spinale in FDI potesse rivelare il loro expertise nell'integrazione dell'impugnare (stimolo tattile) e del sentire (il suono di) una tromba. Dato l'esiguo campione, i nostri risultati vanno considerati come provvisori, pur mostrando una tendenza verso una minore eccitabilità cortico-spinale nei musicisti, rispetto ai non musicisti, mentre si ascolta del rumore bianco, piuttosto che una tromba, indipendentemente dallo strumento impugnato. La minore sensibilità al rumore bianco mostrata dai musicisti può essere presa per un indicatore del loro expertise sensomotorio, ma occorrono ulteriori studi per trarre conclusioni più solide circa le competenze multisensoriali di entrambi i gruppi.

Nel quinto capitolo è stato ideato un paradigma basato sul canto hocketus, nel quale la qualità espressiva di coppie di musicisti è stata testata per mezzo di un metodo di analisi Bayesiano. Tale metodo ci ha permesso di esplorare il timing della coppia senza assumere una stazionarietà della performance, nell'ipotesi che modelli predittivi operino un continuo aggiornamento nel minimizzare gli errori del timing, comparando gli intervalli reali tra due attacchi delle note di ciascun musicista con gli intervalli previsti. Degno di nota è il fatto che il movimento mostrava un effetto diverso a seconda dell'expertise, nel senso che nella condizione dinamica, rispetto a quella statica, mentre tutte le coppie tendevano a essere più precise nel micro-timing (in quelli che abbiamo chiamato errori di "fluttuazione"), le coppie di minore qualità interrompevano molto più spesso delle altre la performance (compivano molti più errori di "collasso"). Sul piano soggettivo, i cantanti hanno riportato sensazioni di agency "condivisa", rispetto a una "we-agency", rivelando un controllo di solo una parte dell'azione congiunta, piuttosto che sentirsi una singola entità col proprio partner nel compimento di quella data azione musicale. Ad ogni modo, questo parametro soggettivo si è dimostrato correlato al parametro oggettivo individuato dalla nostra analisi Bayesiano, ovvero gli errori di durata (fluttuazione, narrazione, collasso), dimostrando la sua plausibilità come indicatore (soggettivo) di un'interazione espressiva. Ancora più altamente correlate con gli indicatori della qualità di un'interazione basati sul timing, rispetto alla agency congiunta, sono risultate le auto-annotazioni di quella stessa qualità realizzate dai musicisti.

Nell'ultimo capitolo discuto i risultati dei nostri studi, rispondendo alle questioni aperte poste nell'Introduzione e proponendo qualche domanda per la ricerca futura. Viene abbozzato poi un modello della musica (e della musicalità) come linguaggio incorporato, allo scopo di integrare approccio predittivo e incorporato.

List of publications and contribution

1) Dell'Anna A., Fossataro C., Burin D., Bruno V., Salatino A., Garbarini F., Pia L., Ricci R., Leman M. & Berti A. (2018) Entrainment beyond embodiment, *Neuropsychologia* 119, 233–240

My main contribution is: co-development of experimental design, execution of the experiment, interpretation of the data and writing the paper.

2) Dell'Anna A., Rosso M., Bruno V., Garbarini F., Leman M & Berti A. (forthcoming), Does a musical interaction in a jazz duet modulates peripersonal space? *Psychological Research*

My main contribution is: development of experimental design, execution of the experiment, co-analysis and interpretation of the data and writing the paper.

3) Dell'Anna A., Sarasso P., Serra H., Ricci R. & Berti (in preparation) Touching the music. Integrating touch and sound of a trumpet. A TMS study

My main contribution is: development of experimental design, execution of the experiment, analysis and interpretation of the data and writing the paper.

4) Dell'Anna A., Buhmann J., Six J., Maes PJ. & Leman M. (forthcoming) Timing markers of interaction quality during semi-hocket singing. *Frontiers in Neuroscience*

My main contribution is: development of experimental design, execution of the experiment, interpretation of the data and writing the paper.

Chapter 1. Introduction

'The players in a jazz trio, when improvising, are immersed in just such a web of causal complexity. Each member's playing is continually responsive to the others' and at the same time exerts its own modulatory force'. Andy Clark (1997, p.165).

1.1. Interaction

As a (jazz) musician and (experimental) philosopher, I have found Clark's description inspiring since the beginning of the present study, whose focus is embodied musical interaction in dyads of both musicians and non musicians. While Clark used this description to sum up his view of embodied mind, that is, as a metaphor of the interplay between brain, body and world necessary for a thorough understanding of mind and cognition, in this series of studies I actually tried to identify several markers of embodied interaction in different musical experimental contexts.

Joint action has been extensively investigated in cognitive science for more than a decade. A working definition put forward by Sebanz et al. (2006: 70) states that a **joint action** is "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment". While lifting an object together has been a rather widely studied instance of joint action (Marsh et al. 2009), the change in the environment mentioned in the above definition may be at the same time subtler and deeper, as when, for instance, two persons exchange gazes in order to read each other's intentions (Becchio et al. 2018). Actually, the need for comprehension of social interaction has been recently reiterated by a number of neuroscientists, stressing that the social mode is arguably the default mode of homo sapiens' brain, not to mention other social species and mammals in general (Caccioppo et al. 2010, Schillbach et al. 2013, Hari et al. 2015). Therefore, it is urged that brain studies develop appropriate methodologies to deal not only with action observation (like classic mirror neurons paradigms), but also with contexts in which two or more subjects modulate each other's behaviour on the fly, be it for competition or cooperation. It is well known that the mirror neurons system is a brain network that is recruited similarly during action perception and production (Rizzolatti & Sinigaglia

2010). An early suggestive finding in order to overcome the limits of a "spectatorial" paradigm (Reddy & Uithol 2015) came from Newman-Norlund et al. (2007), who showed higher BOLD activation in fronto-parietal areas (which are supposed to match the human mirror neurons system) during complementary, rather than imitative, action planning (of a power or a precision grip of an object). These authors found that the very same neural network responsible for passive understanding of observed actions is active (indeed, it is more active) in (preparing) a possible interaction. Mother-infant exchanges epitomize the essence of social interaction, in that observation is always embedded in the dynamic processes of adaptation, reaction, incitement etc., well before any conscious awareness of the context from the infant side, portraying what De Jaegher & Di Paolo (2007) call "participatory sense-making". Hyper-scanning, the simultaneous acquisition of cerebral data from two or more subjects, provides an interesting possibility to explore social interaction, since it takes into account more than one individual at the same time, although the results imply interpretations that are far from straightforward (Konvalinka & Roepstorff 2012, Babiloni & Astolfi 2014, Hari et al. 2015). In the present series of studies, some of the previous concepts (e.g. joint action, participatory sense-making, social mode) will be assessed and exploited in a couple of brain and behavioural experiments involving real time interaction in dyads of both musicians and non-musicians.

1.2 Musical interaction

Ensemble music is a sophisticated form of joint action that, perhaps not surprisingly, has allowed for about a decade a balanced study between controlled experimental conditions, on the one hand, and ecologically valid setting, on the other hand (D'Ausilio et al. 2015). According to Keller's model (2008, Philips-Silver & Keller 2012), interpersonal coordination in a music ensemble relies on a combination of higher-order cognitive processes, like sharing a global idea of the musical composition at stake (which, in turn, depends on socio-cultural conventions), and lower-order cognitive-motor competences, like **mutual adaptive timing**, prioritized integrative attending and anticipatory imagery (Figure 1). These processes may somehow characterize every kind of joint action (Vesper et al. 2010), but in a musical context they amount to the fact that:

1) two or more subjects need their temporal playing coordination be so tight and flexible to cope with unintentional micro-perturbation of timing, due to the intrinsic variability of human actions, and, on the other hand, with intentional timing variations due to expressive purposes (accelerando/ritardando). Phase and period correction are two mechanisms put forward to explain such competences (Repp & Sue 2013).
2) A musician needs to pay attention not only to what he is playing, but also to what the ensemble is playing, prioritizing his resources for the former process, without loosing track of the latter. Internal time-keepers have been postulated to keep track of the multi-layered structure of ensemble music, being it quite often composed of rhythmic sections, intertwined melodic lines and, more generally, different parts according to the performance/composition (London 2004).
3) Musicians need to anticipate their partners' playing to some extent, if they aim at keeping their performance stable and coherent. Keller & Appel (2010) demonstrated, for example, that the most synchronized among a few piano duets were those formed by pianists with higher imagery vividness in a task of notes continuation without auditory feedback. Also adaptive timing and attention may depend on such a skill insofar as they are likely based on internal models governing individual and joint actions (see below).

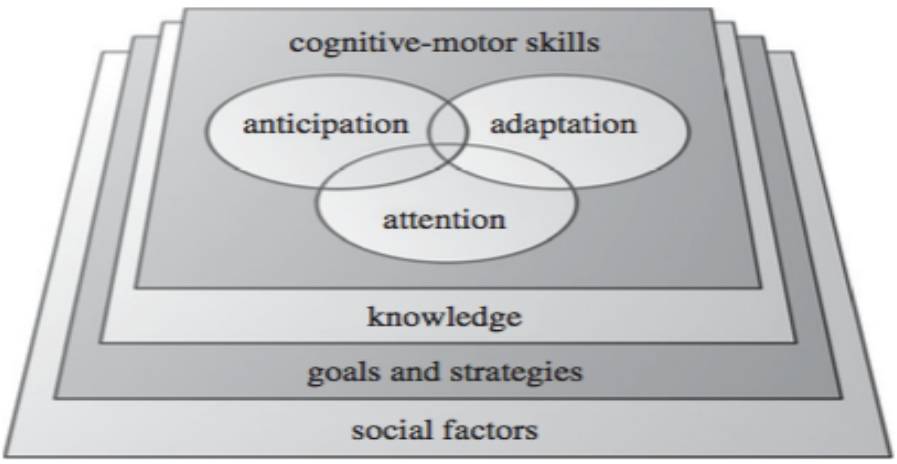


Figure 1. Keller et al. (2014)'s model of the competences needed to reach interpersonal coordination in a music ensemble.

An emerging and promising theoretical framework apparently able to unify all the previous three aspects is the so-called "**predictive coding**" approach (Friston 2013, Clark 2016), whose main aim is to investigate the brain as a predictive machine, as already von Helmholtz and cybernetics did, but making intense use of Bayesian inference as a unifying functional/computational principle. The circular sensorimotor causality this inferential process consists of means that "external states cause changes in internal states, via sensory states, while the internal states couple back to the external states through active states—such that internal and external states cause each other in a reciprocal fashion. This circular causality may be a fundamental and ubiquitous causal architecture for self-organization" (Friston 2013, pp. 2–3). Music perception and production are particularly suited for a predictive coding approach, given the intrinsically hierarchical structure of music from both the melodic (cells within phrases within sections) and rhythmic (beats within cells within meters) viewpoint (Salimpoor et al. 2015, Koelsch et al. 2019), but also musical interaction can profit from such a framework. Indeed, the sensorimotor loops necessary for an individual action to take place, predicting the outcome of a given action and adjusting it in case of wrong sensory feedback, can be translated into social terms (Wolpert et al. 2003, Friston & Frith 2015, Volpe et al. 2016, Brattico & Vuust 2017). In the latter case, we may predict the consequences of an action of ours on a partner (say, accepting to be kissed), while the sensory feedback would be provided by the partner's reaction (say, avoiding us), which, in turn, allows for an adjustment of our action (say, pretending to reach something just behind the partner) to minimize prediction errors.

In a musical context, let this action be the attack of the theme after three introductory measures of the jazz standard *Autumn Leaves*. The musician who plays the theme must adapt to the tempo set by the rhythm section (say, piano, bass and drums), attending to its own sound without neglecting the others' and predicting their correct unfolding. After playing the first two notes the soloist realizes that neither the bass nor the piano changed the chord leading to the real first measure of the tune, therefore he adjusts his trajectory, turning those two notes in a sort of ornament preceding the theme, whose beginning is postponed for a measure. It is worthwhile to stress that such processes need not be fully aware, since internal models are supposed to work in a nested hierarchy, from very low levels (close to reflexes) to conscious levels (very close to propositional thought, see Friston & Frith 2015).

Oddly enough, Philips-Silver & Keller subsume without further explanation the three above-mentioned sensorimotor processes required to play ensemble music under the category of (interpersonal) **entrainment**, defined as "the spatiotemporal coordination of two or more individuals, often in response to a rhythmic signal" (2012: 3). Now, as it seems already clear by this definition, the concept of entrainment would be suitable to include only the first of those processes, that is, mutual adaptive timing (MAT). Indeed, entrainment defines a wide spectrum of dynamics of attraction, ranging from oscillating pendulums, on one extreme, to group coordination in dance or playing, on the other extreme. If entrainment phenomena have to be treated in terms of phase relationships, as most of the existing studies do, adaptive timing is the ideal candidate, along with other essentially temporal aspects of music, like rhythm and meter (Clayton 2013, Moens & Leman 2015). For example, the transitions from higher to lower and then again higher synchronization moments in a jazz trio have been found to correlate with higher groove feelings on the musicians' side (Figure 2. Doffman 2008).

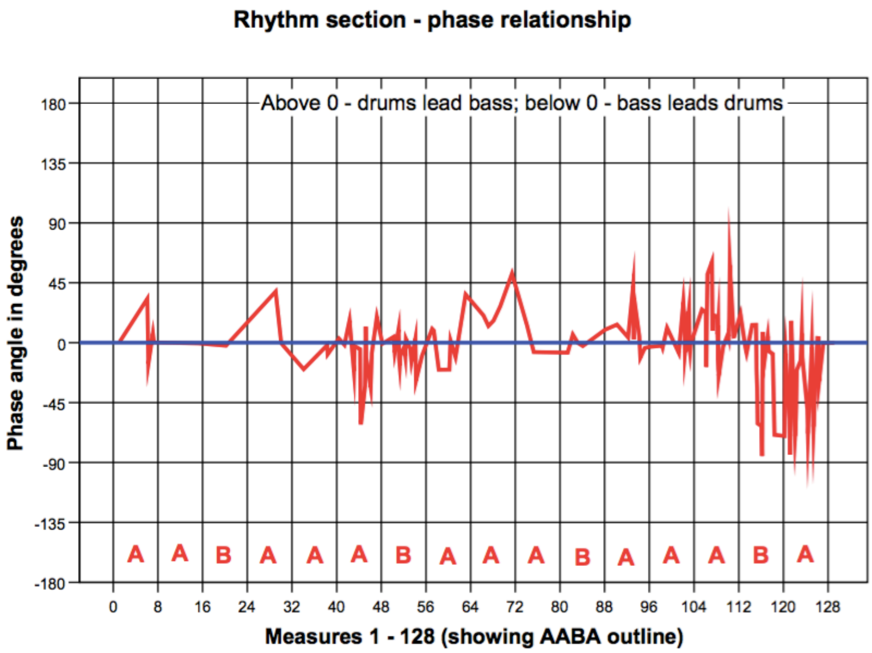


Figure 2. Phase relationships between the onsets of the drums and bass of a jazz trio across the whole duration of a tune (Doffman 2008).

While it is not easy to see how the processes of attention and imagery may be categorized as forms of entrainment, they can be aptly conceived of as internal sensorimotor models in the job of predicting the partner (or the group)'s behaviour, in line with Bayesian approaches. Contrary to Philips-Silver et al. (2011), who seem to be aware of this risk (here dealing only with timing as a case of entrainment), Philips-Silver & Keller (2012) risk to confuse two kinds of explanation. If we frame entrainment in the context of dynamical systems (see Kelso 1995 for a summation), its explanatory powers are to be expressed in terms of general laws, that is, in terms of (differential) equations describing a given phenomenon of attraction. On the other hand, if we are interested in the neuro-cognitive processes at its bases we need a mechanistic account (Colling & Williamson 2014). Even though both styles of explanation have their unquestionable merits and can be fruitfully combined, they should not be confused. As we will see, two of our studies explore adaptive timing in search of ensemble proto-musical competences in non-musicians and how they are modulated by the embodiment of a partner's hand (Chapter 2), or looking for dynamic markers of a singing performance quality that matches subjective reports about joint agency and about that very quality (Chapter 5).

1.3 Embodied cognition

A central ganglion for the present work is embodied cognition, a multi-faceted theoretical paradigm that has been questioning the basic tenets of traditional cognitive science for three decades, in particular, the computational-representational nature of the mind (Varela et al. 1991, Clark 1997, Thompson & Varela 2001, Noë 2004, Chemero 2009, Gallagher 2017). On the contrary, embodied cognition stresses the importance of taking into account the entanglement of body, environmental and social components or the so-called 4E, that is, the embodied, embedded, extended and enactive components of mind and cognition (Newen et al. 2018). It is beyond the scopes of the present thesis to elaborate on each of these aspects, but some of them need explanation in view of the theoretical effort informing our work in combining two apparently opposed frameworks, that is, embodied mind and predictive coding approach.

Introducing the body into the picture entails that mind and cognition are no more conceived of as building representations of the external world by means of neuronal computations, but rather as guiding processes of actions in/on the world, including parts of the world that are particularly meaningful for humans (and animals in general): conspecifics. The previous sentence highlights two of the four E we have to consider in more details, the **embodied** and the **extended** nature of the mind. As to the former feature, we may focus our attention on what Hurley (1998) calls "the sandwich view", according to which the core of the mind lies between perception and action, that is, in those computational processes occurring after the sensory stimuli, but before the motor responses. Embodied approaches have challenged this view, pointing out that what an organism perceives is a function of how it moves and, vice versa, how an organism moves is a function of what it perceives (as Merlau-Ponty and Gibson already put it). Such sensorimotor loops strikingly resemble the "circular sensorimotor causality" Friston (2013) points at in presenting his Bayesian approach and, indeed, they can be further characterized as **active** inferences, in that the whole body, rather than the brain alone, actively enables the inferential (predictive) process. Thus, if the inferential mechanisms are read in sensorimotor, rather than computational-representational terms, the Bayesian approach can coexist and enrich the embodied approach (see Gallagher & Allen 2016 for a similar synthesis proposal and our Conclusions).

The "extended" component of the embodied framework is typically one of the most controversial among the 4E (Menary 2010), since, in its strong version, it implies the inclusion of (parts of) the external world in the computational machinery an organism makes use of to solve a given cognitive problem (Clark & Chalmers 1998' "parity principle"). Tool use is the classic example. When a blind cane user touches the edge of a building in order to orient himself and turn in the right direction, the cane becomes part of his body, as if his own fingers were sampling the environment. Now, consider a joint action like cycling together in a tandem bike. Not only is that action impossible for only one person, but the degree of synchronization necessary to accomplish it is so tight that a kind of "super-ordinate" system may emerge from the coordinated individual actions, an extended cognitive system made up by two (or more) interactive agents. These are two ways of extending the mind, by means of tool use, in the former case, and by means of coordination with a conspecific, in the latter. Whatever the philosophical arguments to include such extensions in the computational machinery of the mind, the previous phenomena, tool use in particular, have been thoroughly investigated in recent cognitive neuroscience and will be briefly presented in what follows.

1) Which cognitive processes is tool use supposed to extend? Although philosophers have pointed also at memory and thought, neuroscience has focused mainly on

body and **peri-personal space** perception. Rizzolatti et al. (1981) discovered in the ventral premotor cortex (vPC), putamen and intra-parietal sulcus (IPS) of macaque monkeys visuo-tactile bimodal neurons discharging both when an object appears close to the body and when it touches the body. Insofar as such neurons are body-part centred, codifying for the space of and around the hand, the head or the torso, they may be considered as the neural correlates of the body space (the proprioceptive and tactile space) and the peri-personal space (the multisensory space reachable by the arms). As to the body space, Graziano et al. (1999, 2000) demonstrated that those neurons' receptive fields are activated by objects in the vicinity of a fake hand (while the monkey's real hand is occluded from view) and by the position of the fake hand, after it is embodied by means of a synchronous stimulation of both the fake and the real (occluded) hand. This is a well-known phenomenon called "the rubber hand illusion" (Botvinick & Cohen 1998), in which a fake hand is judged as one's own hand, if it is placed in a position congruent with one's own body and gets synchronously touched along with one's real (occluded) hand by means of a brush. Therefore, body ownership, the feeling that a body part is owned by a given subject, turns out to be modulated by the position, shape, and movement of the fake hand. Similarly, peri-personal space has been shown to be a plastic phenomenon. Iriki et al. (1996), indeed, demonstrated that, after practicing to collect objects with a rake, the visuo-tactile bimodal neurons of the macaque IPS extended their receptive fields to cover the entire length of the rake (Figure 3). In other words, while before practicing with tools such neurons discharged only when a stimulus appeared close to the hand or the shoulder or touched them, after practicing they discharged also for stimuli appearing in the far space, as far as the rake length. Such a remapping of a near space that becomes far has its equivalent in humans. For example, patients suffering from visual neglect after a stroke showed a dissociation of the near and the far space, with the neglect appearing only in the former, as assessed by means of a line bisection task (Berti & Frassinetti 2000). However, if the line bisection was carried out with a stick, rather than with a light-pen, thus extending the arm length, the neglect transferred also to the far space.

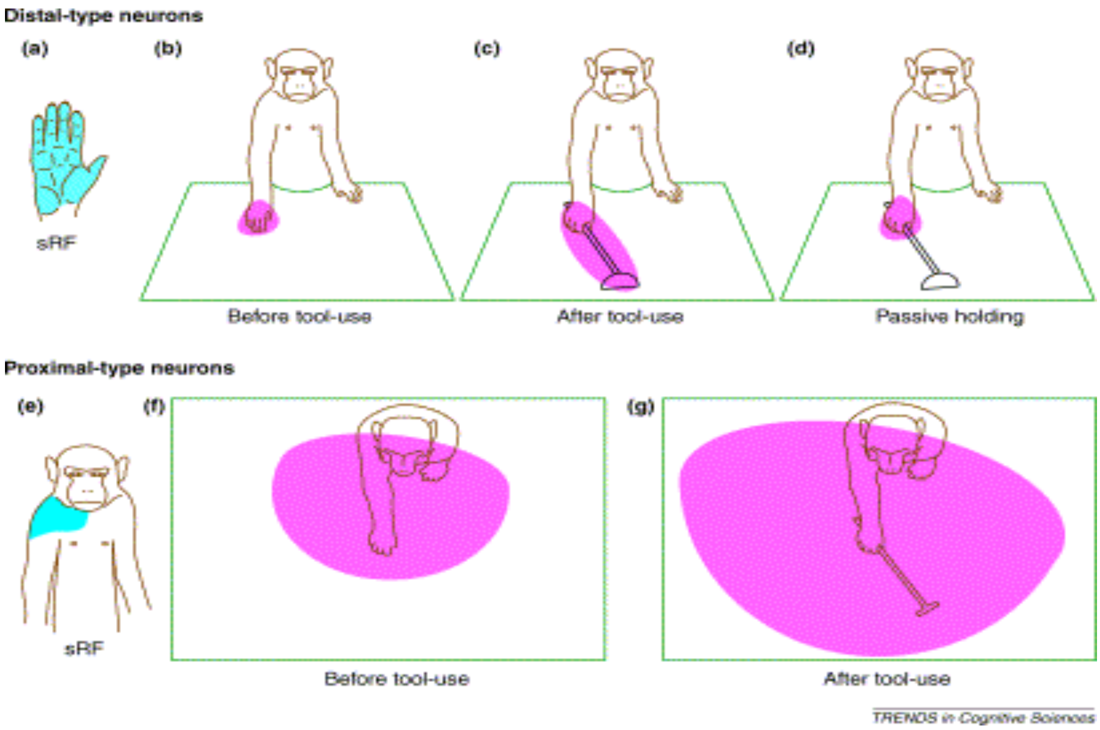


Figure 3. Extension of visuo-tactile receptive fields in bimodal neurons of macaque monkeys' IPS after tool use (Maravita & Iriki 2004).

2) More recently, neuroscience has addressed the possibility that also social interaction has some influence on cognitive processes like body or peri-personal space perception. Soliman et al. (2015) put forward that, during and after a joint action like sawing a candle together with a string, a pair develops a "joint body-schema" that is measurable by means of a visuo-tactile **multisensory integration** (MSI) task. The task consists of a reaction time response to a tactile stimulus delivered either on the thumb or on the index finger while a visual stimulus appears either close to the thumb/index finger of the participant or close to the thumb/index finger of the partner (see Maravita et al. 2002 for the details). Contrary to the solo condition, during the joint condition the incongruity (e.g. thumb touched/index seen) impacted on the reaction times, slowing them down, thereby indicating that an interdependence of the two subjects' body-schema has emerged, due to the joint action just accomplished. Taking advantage of a different MSI paradigm, Teneggi et al. (2013) show that a cooperative, compared to an uncooperative, interaction in an economic game may modulate the peri-personal space of a person in a dyad. Indeed, after the cooperative condition, subjects reacted faster to a tactile stimulus on their hands, not only

when an auditory stimulus was heard close to them, but also when the sound came from a further distance, close to the cooperative partner (see Canzonieri et al. 2012 for the details). Since a response to a tactile stimulus is facilitated by an auditory stimulus in the peri-personal space, thanks to the above-mentioned bimodal neurons, this result is taken as evidence that the peri-personal space got extended, after the cooperative interaction.

At this point it is also worth stressing that Thompson & Varela (2001), two of the main theorists of embodied cognition, already argued that one of the three dimensions of embodiment is inter-subjective interaction (along with what they name "bodily self-regulation" and sensorimotor coupling). As we will see, two of our studies tackle the previous two constitutive "extended" features of the embodied framework, exploring the multisensory space of musicians after a (jazz) cooperative/uncooperative musical interaction (Chapter 3) and in a condition in which they hold their favourite tool, i.e. the musical instrument, while listening to the sound emitted by that very tool (Chapter 4). However, in order to set the stage for each of our single studies, we need to consider how the embodied issues just discussed translate into musical terms.

1.4 Embodied music cognition

A disembodied view typically conceives of music cognition as a computational reconstruction of the hierarchical organization of music in a recursive way, from the basic acoustic stimuli to the wide formal structure of a given composition, much like language cognition (Lerhdal & Jackendoff 1983). Embodied music cognition, on the contrary, takes advantage of the above-mentioned sensorimotor loops as a crucial feature of brain functioning to highlight the role of the body in music perception and production. Firstly, consider again entrainment, the capacity to synchronize a body movement to a rhythmic beat. Such a capacity drives the appreciation of the rhythmic structure of a musical tune by means of a network formed by the posterior parietal lobe, premotor cortex, cerebro-cerebellum, and basal ganglia, giving rise to the phenomenon of "groove" (Janata et al. 2012). Therefore, it can be stated that the same processes that cause bodily motion are involved in rhythm perception. As Todd writes: "If the spatiotemporal form of certain [sensory] stimuli are matched to the dynamics of the motor system, then they may evoke a motion of an internal representation, or motor image, of the corresponding synergetic elements of the musculoskeletal system, *even if the musculoskeletal system itself does not move*" (1999: 120). Iyer (2002), one of the first researchers explicitly working on embodied music cognition, emphasizes that music may evoke different human actions according to its tempo, like breathing, walking and speaking (with frequencies respectively between 0,1 and 1 HZ, between 1 and 3 HZ, between 3 and 10 HZ), but the other way around is also true. Indeed, much existing music compositions lie in this tempo range, suggesting that bodily actions have somehow modelled the way humans create music (van Noorden & Moelants 2002).

Secondly, and related to the previous point, consider how movement can disambiguate a metric structure. In a couple of experiments Phillips-Silver & Trainor (2005, 2007) let infants be passively bounced or adults bend their knees to an ambiguous rhythmic pattern. These subjects' oscillation were set to stress either the second or the third beat, thus rendering either a binary or a ternary meter, as was manifest by their answers afterwards, when asked to recognize which of two different patterns they moved on (while the adults answered verbally, the infants were observed attending to their preferred pattern between those two). Moreover, Su & Pöppel (2012) showed that non-musicians rely more than musicians on their own movement in order to feel the pulse of a rhythmic sequence, missing it when such movements are not allowed. On the contrary, musicians can always rely on their internal clock to understand the sequence even without moving, thus demonstrating the importance of body movement, in particular where expertise is absent. In addition, it is worthwhile to remind that mirror neurons have been shown to depend also on such a sensorimotor expertise. For example, inferior-frontal and parietal areas typically involved in mirror activation, have been found to be more active (in a fMRI scan) in pianists, compared to naïve subjects, while observing piano-playing, compared to non-piano-playing, finger movements (Haslinger et al. 2005).

A framework that might have the resources to hold together the previous empirical findings is Leman's (2007). According to one of its basic tenets: "The human body can be seen as a **biologically designed mediator** that transfers physical energy up to a level of action-oriented meanings, to a mental level in which experiences, values, and intentions form the basic components of music signification. The reverse process is also possible: that the human body transfers an idea, or mental representation, into a material or energetic form" (ibidem: xiii). The physical energy is the acoustic surface of music and the corresponding mental representation is the intention attributed by the listener/producer to that music, "on the basis of a simulation of the perceived action in the subject's own action" (ibidem: 92). In other words, through a repertoire of motor actions (both transitive and intransitive, i.e. gestures), the body maps musical features like rhythm, melodic contours, intensities, tempi etc., promoting their understanding. While, contrary to Leman's intentions, such a formulation may be exposed exactly to the charge of dualism, since a dualism seems to be posed from the start between mental representation and physical energy (Schiavio & Menin 2013), the role of the body in Leman's proposal can be seen differently. Indeed, the mediation the body is supposed to play between a mental and a physical level can be conceived of in a strictly sensorimotor way, suspending any commitment about the nature of what the body is assumed to mediate. Consider again the disambiguation process allowed by moving a body part according to either a binary or a ternary meter on an isochronous pulse. There is no need to attribute physical properties to the sound beats we hear and, on the other hand, mental properties to our subjective experience to the extent that the perception of those sounds is coupled to the body movements necessary to disambiguate them. What counts for an embodied approach to (music) cognition is that exactly such sensorimotor loops, rather than abstract computations, constitutes (music) cognition. Importantly, the sensorimotor mechanism we are dealing with here is twofold. On the one hand, it concerns body morphology, the fact that the human body allows for different actions from other animal bodies, for example, as we saw above, synchronizing around specific frequency ranges, according to the motor action involved. On the other hand, sensorimotor mechanisms have a specific neural counterpart, well represented by the mirror network in both humans and monkeys (Rizzolatti & Sinigaglia 2008 and §1.6 below).

The reference to mirror neurons leads us to clarify how the embodied approach relates to interaction also in the domain of music. Indeed, as we have already hinted at, inter-subjectivity represents a crucial aspect of embodiment (Thompson & Varela 2001), if it is not reduced to internal processes of mindreading or simulation taking place in the interacting subjects' brains. Following De Jaegher & Di Paolo (2007), we could rather talk of "**participatory sense-making**" (see Schiavio & De Jaegher 2017, for a musical application of this concept), pointing to the embodied feature of an interaction provided by the continuous negotiation of spatiotemporal parameters between two (or more) subjects. These authors draw our attention on a very basic joint action like passing together through a door that is too narrow to let two subjects enter at the same time without bending and adjusting to the size and position of each other's body. Note that, if these persons were asked to repeat that action many times, they would likely do it every time in a slightly different manner, thus making it evident that a slightly different dynamics of mutual adjustment unfolded, though resulting in the same outcome (passing through the door). If we apply this scenario to an ensemble music context, some feature emerges, which goes beyond and integrate Keller et al. (2014)'s model described above. Indeed, during a performance, not only temporal, attention and imagery-driven processes are in place, but also are active processes more tightly related to the body and the space between the bodies (see Chapter 3). In Walton et al. (2015), for example, the forearm and head movements of two pianists improvising either on a drone track (a uniform alternation of two chords) or on an ostinato track (a complex four chords progression) were recorded by a motion capture system (Figure 4).

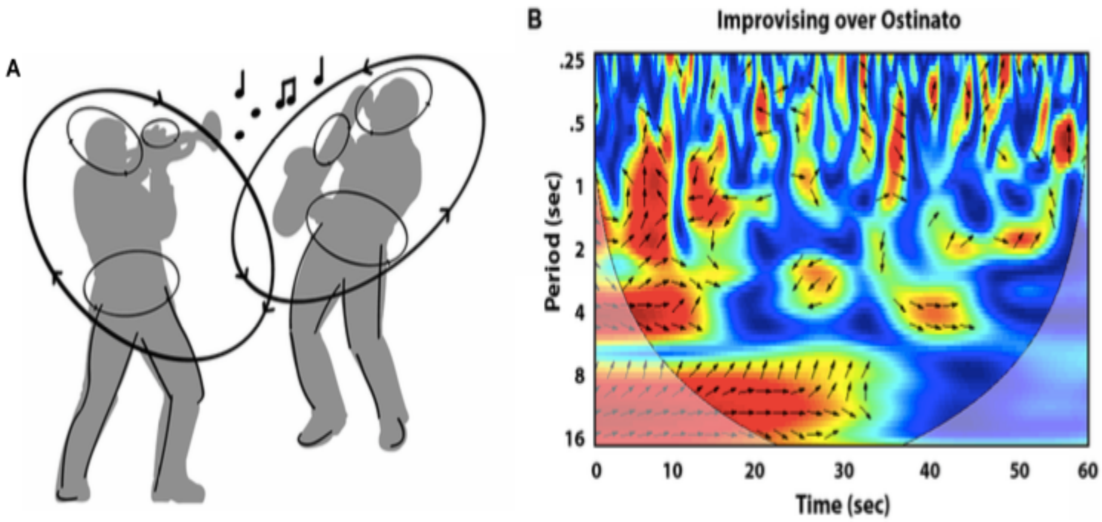


Figure 4. A. Musically interacting bodies. Rings around them identify their potential peri-personal spaces. **B.** CWT plot of two musicians' forearms movement, displaying coherence (red=high, blue=low coherence) at different periods (Walton et al. 2015).

Thanks to cross wavelet transform (CWT), the time series of these movements revealed different periodicities, according to the features of the musical track. The ostinato track, indeed, repeated every 4 seconds, allowing the musicians' movements to coordinate, as it turned out, at multiples of 4 seconds. On the contrary, the drone track didn’t show any specific periodicity, probably due to its simpler structure, that offered the musicians more variety of motion (and, as a consequence, of musical possibilities, but the opposite is also true). The authors drew the conclusion that expressive interactions are guided not only by brain processes, but also by bodily dynamics emerging on the fly, in accordance with one of the tenets of the embodied approach to cognition (see also Walton et al. 2018).

1.5 Expressive interaction

Refining his theoretical framework, Leman (2016) invites us to consider every kind of interaction with music (be it listening or playing, be it alone or in group) as constituted by a **cognitive-motivational loop** that realizes empowerment and reward in the subjects involved in it. We will see in the Conclusion of this thesis whether and how his model could be integrated with Keller et al. (2014)’ s, but now we need to elaborate on the most relevant (for our purposes) of its components, that is, the pro-social orientation induced by the sense of agency (induced, in its turn, by the sensorimotor predictions inherent in) interacting with music. The **sense of agency**, a widely studied phenomenon in the cognitive neurosciences, is the feeling of control of a given person on a given action he/she is accomplishing (Haggard & Eitam 2015). In everyday life it is an implicit feeling, which becomes manifest if something goes wrong, as when you are on the point of pressing a light switch, but the light turns on the instant before you press it: it is not you, who turned on the light, but someone else, hence a weak (or totally absent) sense of agency. On the other hand, being probably built on the prediction of our action consequences, rather than on their real sensory consequences (Berti & Pia 2006), an illusory sense of agency may also ensue. According to Leman (2016), sensorimotor predictions (based on the above-mentioned Bayesian inferences) are able to induce the feeling that a given musical pattern has been produced by a motor action of ours (Leman also reminds the similarity of this idea to Hume’s concept of causality). Such a feeling would be (consciously) illusory in cases of moving to the music without playing it, as in running, dancing, or even simply tapping to the music, but it would be veridical whenever we are really playing the music. Nevertheless, in both cases a rewarding and empowering effect would ensue, due also to a pro-social element that (at least partly) explains the expressive power of musical interactions. This idea is consistent with accounts that emphasize the capacity music exhibits of making people being (Overy & Molnar-Szakacs 2009) or keeping (McNeill 1995) together in time, developing a **joint sense of agency**, a concept on which the philosopher Pacherie (2012, 2014) has recently investigated (see Chapter 5). Arguably, what is still missing from a theory like Leman's (as from many other proposals in the neuroscience and musicology literature) is a more detailed characterization of the relationship between expressive quality and pro-social aspects in music interaction. Insights toward such a link could be found in Overy & Molnar-Szakacs (2009)' Shared Affective Motion Experience (SAME) model, which "suggests that musical sound is perceived not only in terms of the auditory signal, but also in terms of the intentional, hierarchically organized sequences of expressive motor acts behind the signal" (ibidem: 492). Not surprisingly, these authors invoke the recruitment of the mirror neurons network as the neural implementation of such experiences with music. Furthermore, they employ the concept of "sense of agency" (differently from the standard use) to stress the sense of human interaction lying at the core of musical experience, "a sense of the presence of another person, their actions and their affective states" (ibidem: 494, see also Clarke 2005). The idea that a person is lurking behind a musical sound leads to the possibility to conceive of music as an **embodied language**, or, in other words, a technology of group formation and cohesion (Freeman 2000, Cross 2014). This idea resonates with Leman's proposal, when he claims "that musical expression is more than just a habit or settled practice. Expression locks into the biology of human social interaction behaviour, where it is easily linked up with affective states and attitudes" (2016: 49). Unlike natural language, music allows to coordinate in real time behaviours of big size groups, as epitomized in stadium choirs or in war and work songs, and it is well known, particularly in ethnological studies, how such behaviours enhance collective identities, that is, cultural membership (Nettl 2005, Clarke et al. 2015). Admittedly, **musicality**, rather than music itself, should deserve the status of embodied language insofar as it precedes and founds music. "Musicality in all its complexity can be defined as a natural, spontaneously developing set of traits based on and constrained by our cognitive and biological system. Music in all its variety can be defined as a social and cultural construct based on that very musicality" (Honing et al. 2015, p. 2). If biological traits of musicality are likely met in tonal encoding of pitch, beat perception and metrical encoding of rhythm (ibidem), we may think they underlie the communicative character of musicality and, as a consequence, of music as an embodied interactive communicative process (Mithen 2005, Malloch & Trevarthen 2009). Therefore, we may expect to find these traits uniformly distributed among humans, no matter how musically expert they are, representing the prerequisites for musical expertise, rather than its outcome (Mehr et al. 2019).

1.6 Neural sensorimotor underpinnings of musical interaction

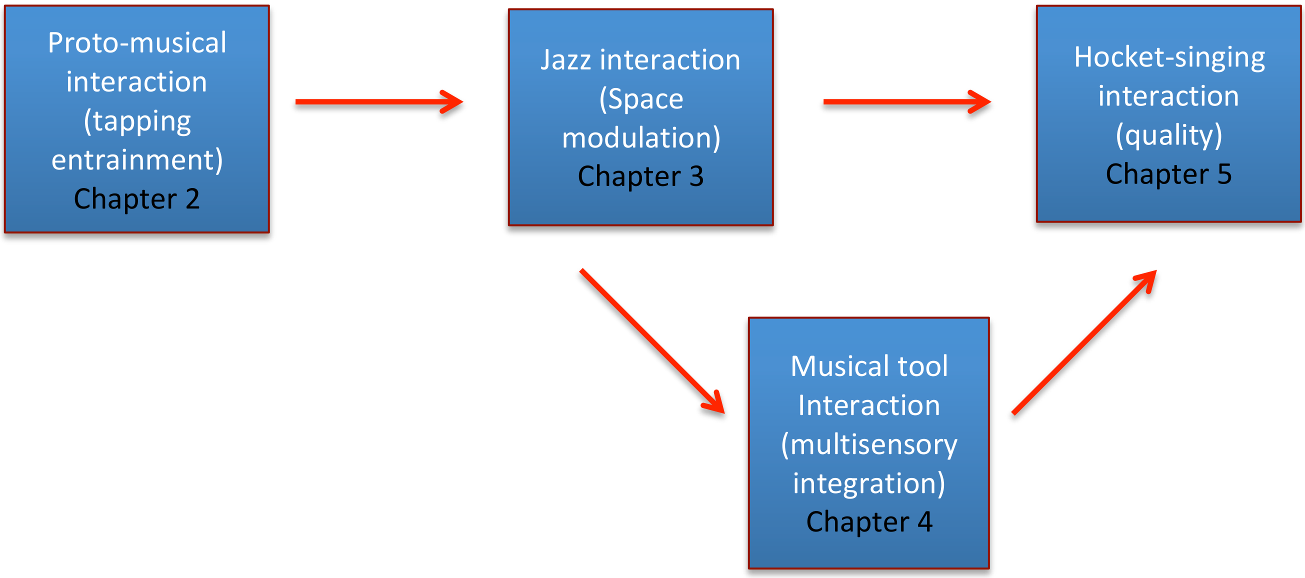
In the last decade, a number of studies have investigated the neural networks enabling musical joint action. Although our work only touches on such aspects, mainly focusing on cortico-spinal excitability by means of single-pulse transcranial magnetic stimulation (TMS) on primary motor cortex (M1), it is worth providing a very short review of the recent acquisitions. The mirror neurons literature has quite extensively demonstrated the involvement of M1 in action observation since Fadiga et al. (1995) pioneering research, in which a subject cortico-spinal activation was shown to enhance while looking at a transitive motor action (grasping) of another person over an object, compared to simply looking at that object (see Aziz-Zadeh et al. 2004 for an auditory counterpart). At least two studies have corroborated such finding in the ensemble music domain. In the first one, Novembre et al. (2012) let a sample of pianists rehearse a couple of compositions before asking them to play their melody part with their right hand either alone or while the left hand part was being played by a (hidden) partner. Single-pulse TMS on the inactive left hand/arm M1 showed higher motor-evoked potentials (MEPs) in the ensemble condition, highlighting that motor representations may arise in response to potential social interaction. In a second study, Novembre et al. (2014) tested another sample of pianists, half of which had rehearsed and the other half of which had not rehearsed a given tune. When asked to adapt the tempo of their right hand to the gradually changing tempo of the left hand played by a partner, after double-pulse TMS on left hand M1, the latter group showed higher accuracy than the former. That is, double-pulse TMS disturbed only those processes relying on the sensorimotor simulation of the rehearsed part played by the partner, a mechanism that is clearly recruited in the real-time coordination of actions generated by the self and the partner.

As we said, being musicality a universal feature of human nature/culture, we do find sensorimotor mechanisms as the ones just described also in non-musicians. For example, Gordon et al. (2018) recently found cortico-spinal facilitation in non-musicians

FDI while they were looking at a three-note piano sequence with sound lagging 200 ms behind the video, compared to a correct audio-visual stimulus condition or to the correct unimodal conditions (either visual or auditory). The authors conclude that sensorimotor predictive models are here at stake, rather than simulation-like mechanisms, given that only the violation of the expected sensory outcome caused an increase in cortico-spinal excitability. On the other hand, if non-musicians are trained to execute simple melodies at the piano, when listening to those melodies their FDI cortico-spinal excitability increases even some milliseconds before the tone onset, thus showing the difference a motor training makes, compared to simply listening (Stephane et al. 2018). This finding is consistent with Candidi et al. (2012), who demonstrated that pianists manually trained with a given composition exhibited higher fingers MEPs than pianist only visually trained with that composition, whenever an incorrect piano fingering was observed.

We are dealing again with sensorimotor processes that are crucial for an embodied approach to (music) cognition in combination with a Bayesian predictive approach. Multimodal sensorimotor neurons are likely the substrate of such processes, in particular in areas like STG and STS, which respond more strongly to auditory-visual stimuli than to auditory or visual stimuli separately (Beauchamp et al. 2004, see Kohler et al. 2002 for echo neurons). However, as we hinted at in §1.1, more interactive brain research methods are paving the way to overcome some of the constraints that has characterized social neuroscience since the mirror neurons discovery. A pioneering study has been Babiloni et al. (2012), which explored the musical performance of three different saxophone quartets by means of simultaneous EEG, discovering that alpha rhythms in frontal areas (Bas 44/45) are correlated with empathy scores in musicians who are observing their own performance (about musical hyper-scanning see also Osaka et al. 2015 and Pan et al. 2018). But this is just the beginning of a (probably long) new tale.

Figure 5. Overview of the thesis. The flow shows a progression from a basic proto-musical interaction to an expressive interaction, passing through the study of peripersonal space and, after a proper short deviation, tool use.



1.7.1 Open questions

1) If musicality can be conceived of as an embodied language, what kind of collective tasks might have the capacity to let it emerge? And how?
2) What is the role of embodiment of a partner's body part in such joint musical tasks? And how are they related to a physiological parameter like cortico-spinal excitability?
3) Is the pro-social nature of music, based on musicality, observable also in the plasticity of the peri-personal space after a musical interaction?
4) Is multisensory integration driven by musical expertise? Can cortico-spinal excitability reveal it?
5) Can expressive interaction be described in predictive coding terms? In terms that highlight its embodied dynamic nature?
6) Is joint sense of agency reflected in the quality of such an expressive interaction? And what kind of joint agency is here at stake?

1.7.2 Overview of the thesis through its research questions

Given the previous theoretical background and open questions, the following research questions have been posed (points 1,2,3 are developed in Chapter 2, points 4 and 5 in Chapter 3, point 6 in Chapter 4, points 7 and 8 in Chapter 5):

1) Can entrainment, as mutual adaptive timing, characterise non-experts proto-

musical interactions like alternate tapping in dyads?

An easy way to investigate entrainment is tapping, a proto-musical motor action allowing also non-musicians to align a body part movement to the beat of the music. Previous experiments have shown that musicians are able to adapt their timing to the timing of the partner's tapping in anti-phase (Nowicki et al. 2013) and that non musicians are able to do the same in an in-phase tapping task (Konvalinka et al. 2010). Under the assumption of an innate musicality, corroborated by the latter of these experiments, I have shown that also non-musicians are able to entrain to the timing of their partner in an alternate joint tapping task. I used correlation of asynchronies as a method to investigate such a form of entrainment, even if the need for a more dynamic measure emerged, leading to the analysis developed in a later experiment (see point 7).

2) Is such an entrainment modulated by the position of the partner?

The alternate tapping task has been carried out in three conditions: alone with the metronome, with a partner in front of the subject and with a partner beside the subject, in a position congruent with his/her body such that the partner tapped with the left hand, while the subject tapped with the right hand. The latter condition exploits the rubber hand illusion paradigm, in which an alien hand is embodied, that is, it is felt as owned by the subject, given particular constrains (Botvinik & Cohen 1998, Garbarini et al. 2014). The results of my experiment highlight that entrainment overcomes the effects of such a form of embodiment, in that mutual adaptive timing holds both when the partner is in front of and beside the subject (but it does not in the alone condition, i.e. when two subjects' asynchronies in the alone condition are correlated).

3) Is the cortico-spinal activation at the basis of such a motor action modulated by the presence and position of the partner?

When the subject embodies an alien hand cortico-spinal excitability tends to decrease, compared to when he does not (Schutz-Bosbach et al 2006, Della Gatta et al. 2016), as if an interaction context sets the motor system to be engaged, while an embodied partner ('s hand or arm) results in no social interaction. MEPs recording by means of TMS on M1 first dorsal interossus (FDI) confirmed this idea in my proto-musical task. When the tapping subject embodied the partner's arm (as assessed by subjective reports of agency and ownership), cortical excitability did not differ from the alone condition. On the contrary, when the partner tapped in front of the partner, the sociality of the context brought about higher cortico-spinal excitability, in accordance also with the mirror neurons literature (Fadiga et al. 1995, Novembre et al. 2012).

4) Is the peripersonal space of a pair of interacting musicians modulated by the nature of such interaction, be it cooperative or uncooperative?

Peripersonal space, the multisensory, body-part-centred representation of the space immediately surrounding the body, has been recently shown to be sensitive not only to tool use (Iriki et al. 1996, Berti & Frassinetti 2000), but also to social interaction (Patanè et al. 2016, Pellencin et al. 2018). In particular, it has been shown to extend after a cooperative compared to an uncooperative (economic) exchange (Teneggi et al. 2013). I let pairs of musicians play with a partner playing either the correct or the incorrect harmonic sequence of a jazz standard tune, under the hypothesis that only the former condition would have caused an extension of the musicians' peripersonal space. It turned out, by contrast, that only the uncooperative condition impacted on the size of the peripersonal space, making it disappear, as if the subject withdrew from the uncooperative partner.

5) Are musicians better multisensory integrators than non musicians (in the audio-tactile integration task used as proxy for peripersonal space)?

In order to measure peripersonal space I borrowed an audio-tactile integration task devised by Serino et al. (2007, see also Canzonieri et al. 2012). Indeed, there is evidence that a sound occurring close to the subject, compared to a far sound, facilitates reaction times to a co-occurring tactile stimulus. This allowed me to compare my sample of musicians with a sample of non-musicians already tested on the same task at Turin Department of Psychology laboratories. Coherently with a recent finding (Landry & Champoux 2017), I confirmed that musicians are faster than non-musicians in reacting to audio-tactile stimuli, regardless of the distance of the auditory stimulus. This finding imposed to open a side issue with respect to the main topic of my thesis, i.e. interaction. Since musicians are better multisensory integrators than non musicians:

6) Is their cortico-spinal excitability, compared to non-musicians', modulated by the

integration of a touched tool and its corresponding sound (when the tool is a trumpet and the sound is the sound of a trumpet)?

It is well known that tool use can modify not only peripersonal space, but also body-schema (Maravita & Iriki 2004), and a musical instrument is a perfect candidate for embodiment in the musicians' body schema (Nijs 2017). A recent study found that looking at food images while holding eating utensils induced higher cortico-spinal excitability in the masseter muscle compared to holding different tools and/or looking at non-food images (Yamaguchi et al. 2014). Likewise I reasoned that musicians, in particular wind instrumentalists, should show higher cortico-spinal excitability in M1 FDI while holding a trumpet and listening to the sound of a trumpet, compared to different sound-tool combinations (scissors and white noise) and compared to a sample of non-musicians. Current temporary data show a sound-tool interaction, but for both groups and in the opposite direction from our predictions, i.e., a trend towards higher cortico-spinal activation for scissors/trumpet than other conditions. However, an interesting interaction also emerged between group and sound, meaning a trend towards lower cortico-spinal activation in musicians compared to non-musicians while listening to white noise.

7) Can timing expressive quality of a singing dyad be captured by a Bayesian, predictive coding approach?

The analysis carried out in the experiments described so far, although mainly concerning musical interaction, are not interactive in themselves. A central aim of my last experiment was to devise a way to capture the dynamic of a singing dyad in order to assess the expressive quality of a hocket performance, focusing on timing. While the main part of studies on timing in pairs of musicians have used some form of correlation of asynchronies or mean signed asynchronies (Goebl & Palmer 2009, Palmer et al. 2013, Clayton et al. 2018, Heggli et al. 2018), we tried to develop a method that did not assume stationarity, that is, a method that cope with the intrinsic variability of human behaviour, be it musical or not. Given the alternate nature of hocket singing, we chose the inter-onset intervals between two singers' notes (starting from the score the singers had to sing) and computed in Bayesian terms a duration error, which was updated during the performance. This resulted in a dynamic measure of timing quality: the lower the errors, the higher the quality.

8) Is such an objective parameter correlated with subjective assessments of the performance concerning its quality and perceived sense of joint agency?

Since I was also interested in the subjective experience of a musical interaction, I correlated such objective measure of timing quality with self-assessment of the performance quality and feeling of joint agency reported by the singers after the performance. Recently, there has been an intense debate about the concept of joint agency in its two kinds, according to Pacherie (2012): a SHARED and a WE sense of joint agency (Dewey et al. 2014, Bolt & Loehr 2017), the former being the feeling of controlling part of the joint action, the latter being the feeling of blending with the partner in a single entity while accomplishing that action. The way we built our hocket score could have caused a WE-agency, but in fact a SHARED-agency was found. Moreover, we discovered higher correlation for self-annotation than joint agency values with respect to duration errors.

Chapter 2. Entrainment beyond embodiment

Dell'Anna Alessandro [1, 2], Fossataro Carlotta [1*], Burin Dalila [1], Bruno Valentina [1], Salatino Adriana [1], Garbarini Francesca [1], Pia Lorenzo [1], Ricci Raffaella [1], Leman Marc [2], Berti Annamaria[1]

[1] *SAMBA – SpAtial, Motor & Bodily Awareness Research group, Psychology Department, University of Turin (IT)*

[2] *IPEM – Institute for Systematic Musicology, Faculty of Arts and Philosophy, University of Ghent (BE)*

Abstract

Mutual adaptive timing (MAT), the capacity to adapt one's timing to the timing of a partner, is a form of interpersonal entrainment necessary to play music in ensemble. To this respect, two questions can be advanced. First, whether MAT can be seen also in non-musician populations. This might imply interesting theoretical consequences with respect to the hypothesis of an innate inter-subjective musicality. Second, whether subject's MAT can be influenced by the position of the partner's body. This might imply that MAT modulation is guided by changes in the feeling of body ownership and agency, which in turn would affect subject's cortico-spinal excitability patterns. In order to test these hypotheses we employed an alternate joint finger tapping tasks (which can be easily carried out without being expert performers), while single-pulse TMS was delivered on FDI M1. This experimental design allowed us to test MAT in non-musicians and to study cortico-spinal excitability patterns while manipulating partners' body position. Ownership and agency were tested by ad hoc questionnaires. We first found that MAT was present also in a non-musicians population and was not affected by the position of the partner, thus pointing to the universality of such a joint proto-musical competence. Moreover, cortico-spinal excitability was similar when the subject tapped alone ('solo condition') and when the subject tapped with the partner in a position congruent with the subject's body (the 'egocentric condition'). On the contrary, when the subject tapped with the partner placed in front of him (the 'allocentric' condition') cortico-spinal excitability was higher with respect to the solo and egocentric conditions. These results show that, despite the fact that the partner was present both in the egocentric and in the allocentric position, only the allocentric condition was treated as a social ensemble. Interestingly, in the egocentric

condition the partner's body seemed to be treated as the subject's 'own' body. The subjective feeling of ownership and agency were coherent with the physiological data.

Introduction

In everyday life humans can reach highly sophisticated levels of spatio-temporal coordination in order to accomplish a joint-action (Sebanz et al. 2006), as exemplified by two or more individuals playing music or dancing. When such coordination brings forth a rhythmic synchronization between individuals, we can observe the phenomenon of "interpersonal entrainment" (IE) (Philip-Silver et al. 2010, Philip-Silver and Keller 2012, Clayton 2013). While "entrainment" is the dynamic of attraction between two not necessarily animated oscillators (like Huygens' pendulums), IE is a typically human phenomenon (for some limited exceptions see Merker et al., 2008), which may occur more or less voluntarily (Schmidt and Richardson 2008) and is explained, either alternatively or jointly, by dynamical systems theory and mechanistic approaches" (Colling & Williamson 2014, Kaplan & Betchel 2011). As in pendulums, the temporal dimension of IE invokes the notion of "relative phase" between two periodic events: as two pendulums carrying out a number of cycles, particular events in the case of human interactions can be periodic, for example, the relationship between the walking bass and the strikes of the snare drum in a jazz rhythm section (Doffmann 2008). If two such events occur precisely at the same time, then they are in phase (relative phase 0°), if one occurs midway between the other, they are in anti phase (relative phase 180°), but they can also maintain many other ratios, as it is manifest in the huge variety of musical meters (3:4, 5:4, 7:8 and so on) and polyrhythms.

In the rich field of studies on sensorimotor synchronization (see Repp and Sue 2013, for a review) some experiments have been recently run on IE in joint finger tapping, a task that, although implying only a very simple motor act (see Leman et al. 2017, Novembre and Keller 2014), allows for an investigation of the phenomenon also in samples of non-experts. Konvalinka et al. (2010), for example, observed in pairs of non-musicians the capacity to adapt their timing to each other in a finger tapping in-phase task with the metronome, provided that acoustical feedback went in both directions (from one subject to the partner and vice-versa) and from the subject to himself. The authors named "hyperfollower" the unity that emerged from this task. On the contrary, Nowicki et al. (2013) tested a sample of musicians, rather than non-experts, in an alternate tapping task. The choice of an expert sample may be due to the fact that alternate tapping is harder than a synchronous tapping (indeed, in-phase synchronization is more stable than anti-phase synchronization, Repp and Sue 2013). Following Philip-Silver and Keller' (2012) suggestion, we can say that, while synchronous tapping can be attributed to chorusing (a musical joint-action in which individuals make equal contribution, like monophonic and homophonic textures), alternate tapping is a form of turn-taking (a complementary joint-action, like call and response in antiphonae or gospel singing), the latter representing a more complex form of joint-action.

Also Nowicki et al. (2013) found a kind of mutual adaptive timing (MAT) in the pairs of musicians they studied by means of cross-correlations of the temporal series of asynchronies of each partner's tapping relative to the pacing signal, provided that the acoustical feedback went in both directions (while the visual feedback turned out to have a negligible influence). In particular, rather than correcting their partner's asynchronies (compensation), subjects tended to follow them (assimilation), that is, they were late or early relative to the metronome, if their partner was himself late or early. As stressed by the authors: "Members of the (musical) ensemble must coordinate their performance with this basic pulse, as well as with each other's sounds, to achieve a well-synchronized holistic musical interplay" (ibidem). But, and this is our first research question, is such a competence a prerogative of musicians (as a consequence of expertise and exercise) or can it be observed also in non-musicians? If the latter is the case, we might argue that such a form of IE (MAT, by no means the only form) is at the basis of the human rhythmic behaviour, representing a prerequisite rather than an outcome of the musical education, thus strengthening the hypothesis of an innate inter-subjective musicality (Wallin et al. 2000, Levitin 2006, Molloch and Threvarten 2009, Honing et al. 2015, Leman 2016).

The second research question we posed is the following: can a manipulation of the feeling of body ownership (i.e. the sensation that the body or a body part is mine, Blanke et al. 2015, Pia et al. 2016, Garbarini et al. 2014, Garbarini et al. 2015, Fossataro et al. 2016, Fossataro et al. 2017) and agency (i.e. the sensation that a certain action is

accomplished by me, Haggard 2017, Pediemonte et al. 2013, Garbarini et al. 2013) affect the phenomenon of IE? In other words, can IE-MAT be modulated by veridical or non-veridical attribution (to me or to my partner) of the motor act involved in the rhythmic performance?
Body ownership and the sense of agency can be manipulated to a degree that a subject can feel that an external object (and its action) becomes part of his/her own body. One of the most used experimental paradigms that induces this delusion of ownership is the rubber-hand-illusion. Such illusion occurs when a rubber arm is placed in a position congruent with the subject's body and internal with respect to the subject's real hand, which is hidden from view and stimulated with a brush while another brush is touching the rubber hand (Botvinick and Cohen 1998). If the tactile stimulation on the two hands is synchronous, the rubber hand gets embodied after a few seconds, that is, the subject feels as if it has become part of his/her own body and, if it moves, as if the subject is the author of that movement. Schutz-Bosbach et al. (2006) used a paradigm similar to RHI by delivering synchronous or asynchronous visuo-tactile stimulation to the subject's hand and to the co-experimenter's hand. After the RHI procedure, Motor-Evoked Potentials (MEPs) to Transcranial magnetic Stimulation (TMS) were recorded from the right first dorsal interosseus (FDI) muscle during an action-observation paradigm, in which the co-experimenter moved her/his fingers. They found that, after asynchronous stimulation (when the embodiment did not occur), MEP amplitude, registered from the own hand, increased, as it is usually observed in the action observation paradigm (Fadiga et sl., 1995). Indeed, Fadiga et al. (1995) in a seminal paper, using single-pulse TMS on the primary motor cortex (M1), found that cortico-spinal facilitation occurred whenever a subject observed someone acting on an object (e. g. during a grasping action), compared to when he/she simply looked at it. This showed that the observer's motor system immediately activates when another subject is performing a finalised motor act, and in a similar way with respect to when the observer moves himself. Therefore, according to these data in the Schutz-Bosbach experiment (2006), when the experimenter's hand was correctly treated as 'alien', that is as belonging to some other person, the motor system responds in the mirror like fashion, with an increased activity of the cortico-spinal system. On the contrary, after synchronous stimulation (when the experimenter's hand was embodied), identical observed actions, now illusorily attributed to the subject's own body, did not produce any motor facilitation (i.e. the

MEPs amplitude was unchanged with respect to the baseline). The absence of MEP modulation during movement observation following synchronous stimulation can be interpreted as a motor pathways inhibition for own action observation (Ehrsson et al. 2004, Della Gatta et al. 2016). These data show that the motor system has the resources to distinguish between the self and other's body/action (Schutz-Bosbach et al., 2006, but see Decety and Chaminade 2003).

The findings discussed above suggest that when a subject looks at the other's hand movement at least two mechanisms can be activated depending on the ownership ascribed to that hand. Usually, if the observed moving hand is considered to be part of someone else's body, a cortico-spinal facilitation of the own hand is observed due to the mirror neurons system activation (as in Fadiga et al. 1995, and in Schutz-Bosbach et al., 2006). On the contrary, if the other's hand is, under certain manipulations, embodied in the subject's body representation, (as in the RHI and similar paradigm), a cortico-spinal inhibition for the own hand is observed, as if the own hand is disembodied (Ehrsson et al. 2004, Della Gatta et al., 2016). Moreover, as already mentioned, when two (or more) people are involved in the same motor context, a 'joint action' can be pursued and the mirror neuron system is one of the brain networks that activate in joint action context (Masumoto and Inui, 2014).

Novembre et al. (2012), using a musical experimental paradigm, created a joint action context where they let a sample of pianists learn a number of Bach's chorales and afterwards tested them in the following three conditions: participants performed with the right hand the melody alone; they performed the melody with the right hand while a hidden partner was performing the bassline with the left hand (a recording, actually); they performed the melody persuaded that the hidden partner was performing the bassline, but without acoustic feedback. In both joint conditions (with or without sound) the authors found higher cortico-spinal excitability - as indexed by the amplitude of the MEPs recorded from the left FDI, ADM (abductor digiti minimi) and ECR (extensor carpi radialis) - than in the condition in which the pianists played alone. This is, therefore, an example where the motor system seems sensitive to the sociality of the context, activating more complex action plans, which take into account the other as a potential co-actor. The authors conclude that the facilitation effect observed in the joint condition, rather than reflecting a ‘‘copy’’ of the movements associated with the left-hand part, could be taken as a social modulation of the motor system via mirror neuron’ system activation.

To summarize, when two individuals act in the same context, the motor system facilitation/inhibition seems to depend either on the ownership attribution and/or on the sociality of the context. In the first case (ownership attribution) an embodiment mechanism, as that induced by the RHI paradigm, would imply a cortico-spinal inhibition of the own ‘disembodied’ hand, once that the ‘alien’ hand is incorporated. In the second case (sociality of the context), a ‘mirror’ mechanism would be triggered, that implies the activation of the motor system of the observer when a partner is implementing a finalised action. This would entail an increment of MEP’s amplitude in the observer as part of a shared motor situation.

In the present work we took advantage of an alternate joint-tapping task to investigate 1) the capacity of non-musicians to give rise to anti-phase MAT-IE and 2) the possibility that such phenomenon is modulated by the position of the partner’s hand (egocentric vs allocentric position) with respect to the subject: the egocentric position is the one in which embodiment may occur (e.g. Bucchioni et al. 2017) while the allocentric position is the one in which we perceive the body parts of others in every day life (e.g. Fossataro et al. 2016). One important aspect of our experiment is the real interaction it implies, while Novembre et al. (2012)’ set-up had pianists playing with a recording. We first asked subjects to practice, bimanually, alone, alternating tapping with their right and left index finger on two drum pads endowed with a snare and a bass drum sound respectively, at 120 bpm metronome (the preferred tempo for many human movements, van Noorden and Moelants 1999). Such a practice reproduces in a hyper-simplified way Novembre et al. (2012) learning phase of the piano chorales. Afterwards, in order to get a measure of motor system excitation, MEPs were recorded (from the FDI of the left hand at rest) while participants performed the task in the following three conditions: solo, allocentric (they tapped in alternation with the partner, one in front of the other), egocentric (subjects tapped in alternation with the partner, who stayed in a position congruent with the subject’s body). Moreover, subjects had to answer to a Likert-scale questionnaire about the sense of agency and ownership in both joint conditions.

First, if non-musicians are able to assimilate their timing to the timing of the partner, the correlation values of the allocentric condition should be positive and higher than those of the solo condition. As regards to the partner's position, we expected to see higher cortico-spinal excitability when the partner is in the allocentric position, due to mirror mechanisms that the shared action should activate, compared to when the subject taps alone with his/her right index finger in alternation with the metronome. On the contrary, we hypothesised that in the egocentric condition MEPs should be similar to those in the solo condition because the distinction between the self and the other may become weaker, as if there is no longer any partner to interact with. Following the same reasoning, also the behavioural outcome could turn out to be perturbed and MAT-IE could not hold anymore: if I can't distinguish my partner, I won't be able to interact with him in the effective rhythmical ways typical of ensemble music (even in the hyper-simplified way represented by tapping).

Materials and methods

Participants

Twenty right-handed volunteers (13 female, 7 male, mean age = 25.3 years, standard deviation = 5) took part in the experiment. One of them was excluded as outlier in the questionnaire scores. Participants were screened to exclude musical expertise and neurological or medical disease. According to the experimental procedure (see below), subjects acted together with a partner of the same gender (male with male and female with female) to avoid distress in the egocentric condition, since it implied contact between them. The participants did not know each other before and were naıve with regard to the purpose of the study. None of them had history of neurological, major medical or psychiatric disorders and they were free from any contraindication to TMS (Rossi et al. 2009, Bruno, Fossataro and Garbarini 2017). The experiment was approved by the Ethics Committee of the University of Turin and informed written consent was obtained from each participant.

Behavioural recordings

In order to record the mean asynchronies between the tapping of the subjects in the pair and the metronome beats, and then assessing if and which form of entrainment occurred, two circular drum pads (diameter 20 cm) were used linked to an Axoloti circuit board (www.axoloti.com), whose software was specifically programmed to deliver a snare drum sound on one drum pad and a bass drum sound on the other, recording their time stamps at each tap. Subjects could hear the metronome and the sound of each drum pad by means of two headphones. They were sitting on a chair and required to tap in a comfortable way with their right index finger on the drum pad placed on a table in front of them. After a short training phase, in which each of them separately had to tap on both drum pads in alternation with both hands at 120 bpm, the right pad of the subject who got brain stimulated was hidden from his/her sight by means of a cartoon barrier and he/she was asked to tap while looking only at the partner's pad. The tempo of the metronome was always set at 120 bpm and the sound of each pad was cut to last a few milliseconds.

Stimulation and Physiological recordings

Magnetic Stimulation

Motor evoked potentials (MEPs) were elicited by single-pulse transcranial magnetic stimulation (TMS) (Magstim Rapid2; Magstim Co. Ltd, Whitland, UK) with a 70-mm figure-of-eight-shaped coil positioned over the hand area of the right M1. The optimal location for stimulus induction (the location that gave the maximum MEP amplitude) for the left FDI muscle was identified. At this location, the coil was positioned and fixed with the handle pointing backwards at 45 degrees from the midline so as to activate the selected muscle. Then, the resting motor threshold (rMT) was determined as the intensity needed to evoke a MEP in the relaxed muscle of more than 50 μV in 5 out of 10 consecutive trials. The stimulator output was set at 110% of each subject's rMT (56.04%±6.46%, range 46-63% of the maximum stimulator output). Participants who showed a rMT higher than 70% of the stimulator output were excluded from the stimulation phase.

Electromyography recording

Electromyographic (EMG) activity was recorded (MP150, Biopac System, USA), from the left first dorsal interosseous muscle (FDI) by self-adhesive bipolar surface electrodes with active electrode over the muscle belly and the reference electrode over the associated joint or tendon. Signals were amplified and digitalized with a sample rate of 10 kHz, filtered with a band-pass (10 Hz to 500 Hz), and stored in a computer for offline analysis, according to methods used in previous studies (Bucchioni et al. 2016, Fossataro et al. 2017, Bruno et al. 2017).

Task and procedure

The experiment was programmed by using E-prime software V2.0 (Psychology Software Tool Inc., USA) in order to trigger TMS pulses at a controlled timing and trigger the EMG recording. After a short training phase in which the metronome was turned on and the subject was asked to synchronize with it, tapping in alternation with the right index finger on the right drum pad and with the left index finger on the left drum pad, electrodes were placed on the left FDI muscle and the left part of the body was covered with a black cape, in order to prevent the view of the own arm during the experiment. Moreover, we instructed the subject to look at their partner's pad and to start tapping after a pre-recorded voice stressed the first four beats of the metronome, trying to synchronize with the odd beats of it. Then, the experiment started in one of the following three conditions (see figure 1), with the order of conditions randomized across couples:

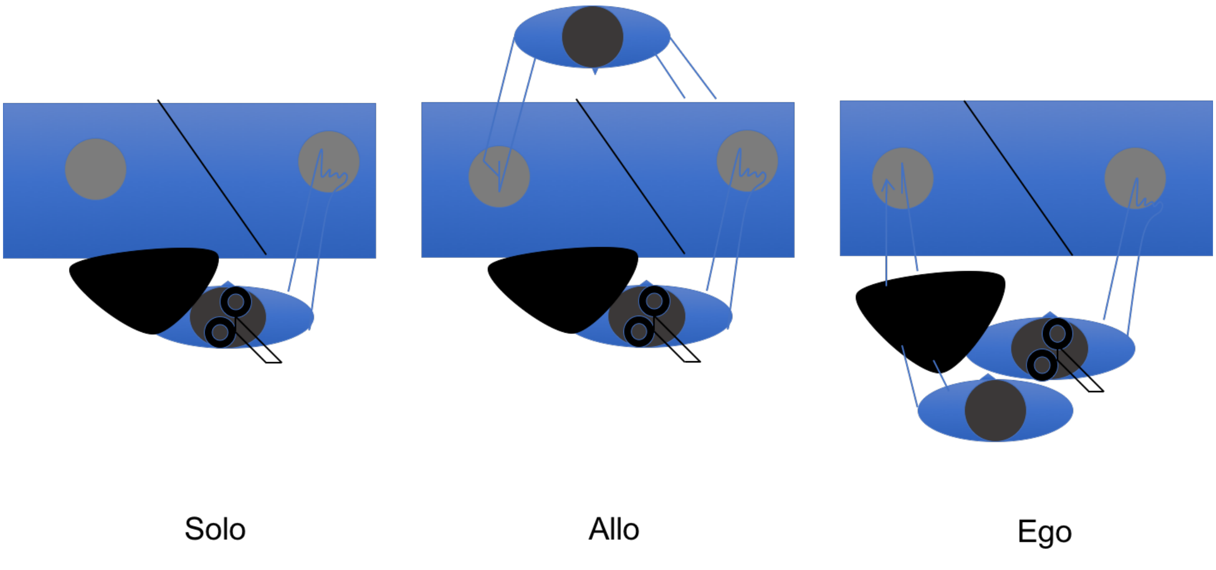


Figure 1. Schema of the three experimental conditions. Solo (left): the subject taps alone with the metronome. Allo (center): the subject taps in alternation with the partner in front of him. Ego (right): the subject taps in alternation with the partner sitting in a position congruent with the subject's body. Single-pulse TMS is delivered on the right

M1 and MEPs recorded from left FDI muscle

1) Solo: the subject taps on his pad with the right index finger on the odd beats of the metronome;
2) Allocentric (Allo): the subject taps with the right index finger on the odd beats of the metronome, while his/her partner sitting in front of him/her taps with the right index finger on the other pad on the even beats;
3) Egocentric (Ego): the subject taps with the right index finger on the odd beats of the metronome, while his/her partner sitting beside him/her taps with the left index finger on the other pad on the even beats. In this condition, the partner taps with his/her left arm placed in a position congruent with the subject's body and covered itself, except for the hand.

In order to check for any corticospinal excitability change related to TMS *per se*, ten baseline MEPs were recorded before (i.e. baseline-pre) and after (i.e. baseline-post) the experimental block, each time the right index finger performed a tapping, with an interstimulus interval of 10 s. The MEPs amplitude recorded during the baseline were used to normalize data recorded during the experimental conditions.
Each experimental condition consisted of six trials of 30 seconds, in which participants were instructed to start on the fifth beat of the metronome and go on until a pre-recorded voice said "stop" (28 seconds later), gathering about 28 time stamps for each subject of the couple (tempo always set at 120 bpm). A 10 seconds inter-trial pause followed. Three TMS single pulses for each trial were delivered online in correspondence of the hypothetical tenth, eighteenth and twenty-fourth tap of the partner, giving a total amount of 3x6=18 MEPs for each condition (plus 20 baseline MEPs).

Once both subjects' motor threshold was established, the experiment took approximately 40 minutes, 20 minutes for each subject who got stimulated. It should be taken into account that sometimes one of the partners could not be brain stimulated, either because of lack of time (e.g. the search for the first subject's threshold took too long) or because his/her threshold was too high. Nevertheless, in order to correlate the time series of the pair, we recorded also the time stamps of those partners who could not be stimulated (figure 2).

Immediately after both the allocentric and the egocentric conditions a Likert-scale questionnaire (−3=strong disagreement; +3=strong agreement; 0=neither agreement nor disagreement) about the sense of agency and ownership was administered. As for agency, the items were: "The hand I was looking at moved exactly as I wanted", "I felt as if I was in control of the movements of the hand I was looking at", "I felt as if I was causing the movements of the hand I was looking at" (these are the real questions, then we added three control items). As for ownership, the items were: "I felt as if I was looking at my own hand", "I felt as if the hand I was looking at was part of my body", "I felt as if the hand I was looking at was mine" (plus three control items).

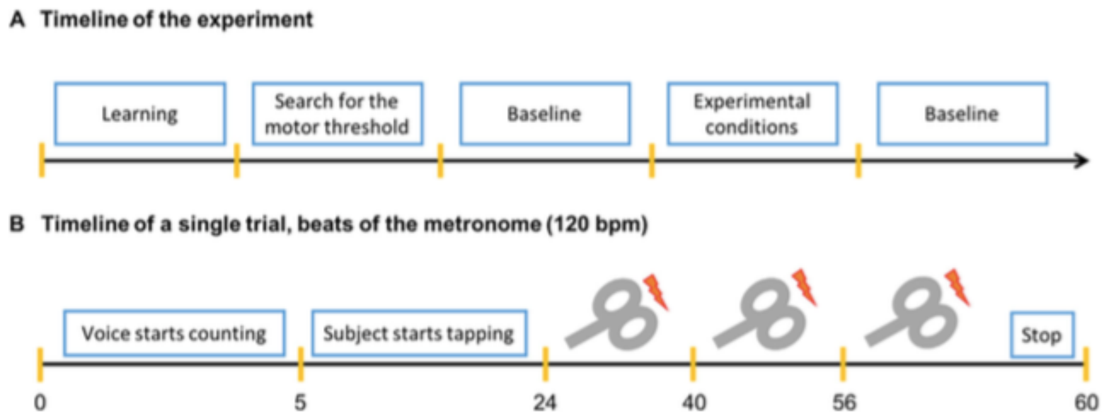


Figure 2. Timeline with the different sections of the experiment (A). Timeline of a single trial (30 seconds) in beats of the metronome at 120 bpm (B).

Data analysis

Behavioural analysis

First, raw asynchronies for each trial were computed by subtracting the onset time of each event in the pacing metronome sequence from the nearest registered tap time. Then we addressed serial dependencies in tap timing by examining cross-correlations of asynchronies in each trial and averaging them: partially following Nowicki et al. (2013), we call "lag 1 auto-correlation" the correlation of the series of asynchronies generated by each individual alone with the same series shifted by one and "joint-correlation" the correlation of the series of asynchronies generated by co-acting members of a dyad. We will not report the results for the former measure in the current article, since we were interested in the social dimension of the task, which is mainly expressed by the latter. Neither we report mean asynchrony measures, since we were interested rather in their correlation as a mark of MAT-IE. Then, joint-correlation was the variable of interest, that is, the correlation of the series of asynchronies generated by the subject with the series of asynchronies generated by the partner. Positive values of the joint-correlation coefficient suggest a greater tendency for temporal assimilation than compensation in mutual adaptive timing, that is, a tendency to follow the direction of the partner's asynchronies with respect to the metronome (late, if the partner is late, early if the partner is early). Assimilation is a form of entrainment, whereas negative values indicate compensation, that is, correction of the partner's asynchronies. In order to have an effective measure of the joint-correlation in the Solo condition (to be compared to Allo and Ego conditions) we correlated the series of asynchronies of each subject of a pair when he/she tapped alone with the partner' series in the same condition. Then, we performed a one-way ANOVA with a within-subjects factors "condition" (three levels: Solo, Allo, Ego) and post hoc comparisons using Bonferroni's test.

As for the questionnaire, the mean value of the three ownership statements and the three agency statements used in the subjective rating questionnaire, in the allocentric and egocentric conditions, was obtained and used as a dependent variable. An outlier was removed and a paired T-test (two tailed) was performed on 19 subjects comparing Allo and Ego condition.

Physiological analysis

EMG data were analyzed offline using AcqKnowlege software (Biopac Systems, Inc., Santa Barbara, CA) to measure the peak-to-peak amplitude (in μV) and MEPs with an amplitude lower than 50 μV were discarded from analysis. Trials showing pre-activity (EMG signal greater than 50 μV) in the time window of 100 ms before the TMS pulse were excluded from analysis. Normal distribution of the residuals was checked using the Shapiro-Wilk test ($p>0.05$) and the appropriate parametric tests were performed by Statistica Software 7 (StatSoft, Inc., Tulsa, UK). The mean MEPs values acquired during baseline-pre were compared to baseline-post by means of a paired T-test (two-tailed). According to the negative results in the baseline analysis, the mean MEPs amplitude of the baseline were used to normalize data of the experimental conditions. For the experimental condition a MEPs ratio (MEP ratio= $MEP_{obtained}/MEP_{baseline}$) was calculated and used as dependent variable in a one-way ANOVA with a within-subjects factors "condition" (three levels: Solo, Allo, Ego). Post hoc comparisons were carried out by means of Bonferroni test.

Results

Behavioural results

The serial dependencies between asynchronies generated by the pairs of non-musicians are plotted in figure 3.

The first thing that can be noticed is that Pearson correlation coefficients (r) for conditions ego (mean±SE=0.421±0.038) and allo (0.424±0.051) are both positive, significantly greater than zero and very similar in magnitude, contrary to the solo condition (0.044±0.028). As we said, we obtained the coefficient in the solo condition by correlating the series of asynchronies generated by each partner separately, when they tapped alone in alternation with the metronome.

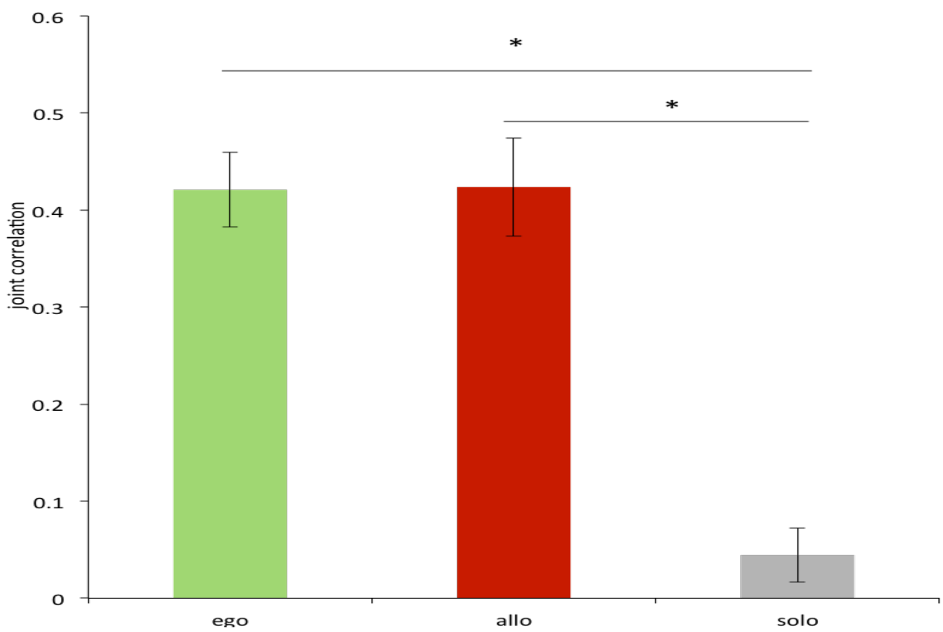


Figure 3. Average joint-correlation of asynchronies for each condition. Error bars represent the standard error of the mean.

The ANOVA found a significant effect of condition [$F(2,36)=34.97$; $p<0.00001$; $\eta^2=0.66$; power=1]. At post hoc comparisons our results show a clear behavioral difference between solo and allo conditions ($p<0.00001$) and between solo and ego conditions ($p<0.00001$), but no difference between allo and ego conditions ($p=1$) (Bonferroni correction). Since a stronger tendency for temporal assimilation than compensation is evident (the correlation value is positive, with a medium effect size), we can conclude that an entrainment in the form of mutual adaptive timing emerged.

As for questionnaires, the ratings of ownership ($t_{(18)}=2.635$; $p=0.017$; $dz=0.61$) and agency ($t_{(18)}= 2.375$; $p=0.029$; $dz=0.55$) of the partner's hand in the egocentric condition were significantly higher than those in the allocentric condition, meaning that some kind of embodiment occurred in the former, but not in the latter condition (figure 4).

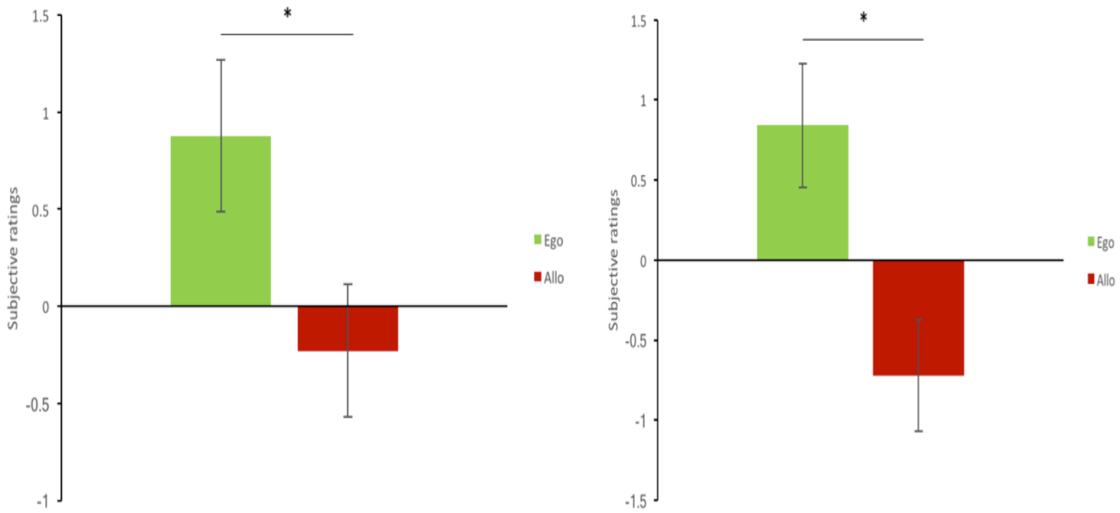


Figure 4. Average score of the questionnaires for the feeling of ownership (A) and agency (B). Error bars represent the standard error of the mean.

Physiological results

In the baseline analysis, the T-test (two-tailed) did not show any significant results ($t_{(18)}=0.34$; $p=0.73$). This suggests that TMS *per se* did not induce any change in cortico-spinal excitability and that the cortical excitability was unchanged from the beginning compared to the end of the experimental block.

The one-way ANOVA found a significant effect of condition [$F(2,36)=5.98$; $p=0.006$; $\eta^2=0.25$; power=0.85], suggesting a different MEPs modulation between conditions. At post hoc comparisons (Bonferroni correction), contrary to the behavioral data, cortico-spinal excitability in allo condition was significantly higher compared to both ego (p=0.023) and solo (p=0.009) conditions. No difference between ego and solo condition was found (p=1). In the plot in figure 5 a striking similarity can be observed between solo and ego conditions, suggesting, along with the answers to the questionnaires, that a form of embodiment occurred in the latter.

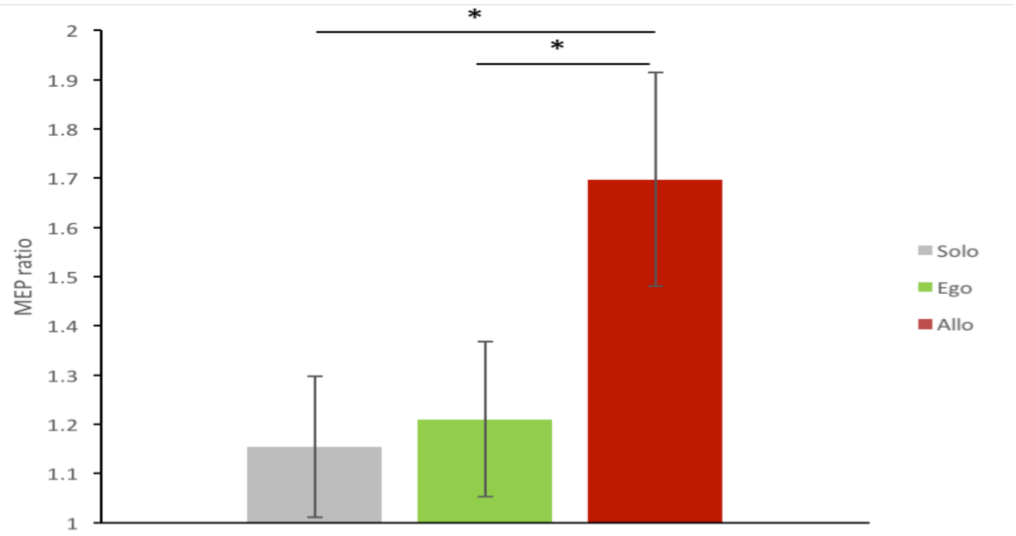


Figure 5. Values of the MEP ratio ($MEP_{obtained}/MEP_{baseline}$) for each condition. Error bars represent the standard error of the mean.

Discussion

In the present study we investigated the phenomenon of interpersonal entrainment (IE) in an alternate joint-tapping task between pairs of non-musicians and its possible modulation according to the spatial position of the subjects in the pair. Moreover, we wanted to see whether the manipulation of the spatial position we employed could elicit embodiment phenomena, similar to those observed in the rubber hand illusion (RHI) paradigm, and whether this could affect both IE and the motor system excitability. Accordingly, we hypothesized that, using single-pulse TMS, different IE and MEPs should be registered, depending on the condition of the experiment. First, we have shown that mutual adaptive timing (MAT), one of the many forms that IE may assume (see Introduction), is not restricted to musicians, but characterizes, at least in our paradigm, also non experts' performance. Moreover, the form of IE that we found was not modulated by the partner's position in the couple, overcoming the embodiment effects due to it. Finally, we have shown that the spatial position of the partner modulates not only the feeling of body ownership and agency (observing an embodiment in the ego condition, Botvinik and Cohen 1998, Bucchioni et al. 2016), but also the physiological value of cortico-spinal excitability, according to whether the partner tapped in front of the brain-stimulated subject (allocentric position) or besides him/her, with his/her left hand in a position congruent with the subject's body (egocentric position). We will discuss each of these points separately.

The form of IE represented by MAT is an essential prerequisite to play in ensemble, along with two other cognitive competences: prioritized integrative attending and anticipatory imagery (Keller 2008). The former is the capacity to pay attention not only to one's own musical part, but also to what the rest of the ensemble is playing, while the latter is the use of internal models to foresee not only what the musician is on the point of playing, but also what the other musicians will do in the short run. Several experiments have shown mutual adaptive timing to occur in musical contexts (Schoegler 2000, Keller et al. 2007, Goebl and Palmer 2009), and some studies (e.g. Nowicki et al. 2013) that used alternate tapping tasks, suggested that only expert musicians would show it. Actually, Nowicki et al. (2013) only tested expert musicians on the possible assumption that anti-phase synchronization would be too difficult for non-expert (see also Repp and Su 2013). Nevertheless, in the present study we found a similar IE effect also in non-musicians. This is a very interesting outcome because, suggesting a universality of IE-MAT, it indicates that it is a prerequisite for playing in ensemble, rather than a result of the long sensorimotor training that every musician has to complete before mastering an instrument, at least in our culture. Leaving aside the rich debate on the possible evolutionary origins of music (Wallin et al. 2000, Honing et al. 2015), we would like to stress that (almost) every society has developed some form of music and that, contrary to modern societies, primitive societies show little difference between music producers and consumers. Therefore, it is not surprising that when non-musicians (who can be considered more consumers than producers) tap in alternation with the pacing signal of the metronome, they are able to adapt their timing to their partner's (see also Koelsch et al. 2000).

Interestingly, this capacity seems to be somehow unconscious since, as phase-error correction (Repp and Su 2013), it happens on a milliseconds timescale and without any explicit instruction for the two subjects to reciprocally synchronize (in fact, subjects were asked to synchronize with the metronome only). This 'automaticity' behind IE-MAT may be a further argument for assuming its 'universality'. Crucially, the IE-MAT was not modulated by the 'embodiment' of the partner's hand in the ego condition. This finding seems again to suggest that IE-MAT is not bonded to the supposed 'agent' of the action, but, instead, is apparently governed by the intrinsic characteristic of the joint proto-musical system, overlooking the mechanisms responsible for the self-other distinction.

As it is a lot more evident in a real musical context (Cross 2014), also in our set-up the regular pulse of the metronome along with the beats produced by tapping on the two drum pads outline a sort of basin of attraction (Leman 2016) around which participants share attention, cognition and action. Several studies have recently stressed how being and keeping together in time (McNeill 1995, Overy and Molnar-Sacks 2009) may induce a sense of affiliation, blurring the self-other distinction (Hove and Risen 2009). Such a capacity for strengthening the social bonding, such "bio-technology of group formation" (Freeman 2000), leads us to consider music, or at least musicality, as an eminently inter-subjective phenomenon. In this respect, the philosopher Elizabeth Pacherie (2012), in her phenomenological analysis of joint-action, has distinguished a SHARED sense of joint-agency from a WE sense of joint-agency: in the former kind, participants in the joint-action have different roles and their actions are complementary, whereas in the latter kind, roles and actions tend to be so similar that the sense of the (acting) self may weaken, in favour of a super-ordinate unity. The IE that we have found also between two non-musicians might be explained by this mechanism of WE sense of joint-agency, whose physiological markers have still to be identified. A clue could be found in Fairhurst et al. (2013), who, using fMRI, assessed an optimal range of synchronization and mutual adaptive timing between a tapping subject and his (virtual) partner, characterized by the activation of the Default Mode Network (cortical midline structures in conjunction with premotor areas), whereas different ranges activated right lateral prefrontal areas associated with central executive control processes. Contrary to the latter, the activation of the former mechanism points toward a fluency of the (proto-musical) interaction and, again, toward a blurring of the self-other distinction. This could be due also to the higher predictability of the optimal synchronization condition. Actually, Bolt & Loehr (2017) recently showed that the rating of SHARED agency in a tone sequence production was higher the more predictable was the partner of the joint action.

Another important finding of our experiment is that the spatial position of the partner's tapping hand seems to modulate both the sense of ownership and the sense of agency in a way similar to that usually found in the RHI paradigm. Indeed, the results of the questionnaires we proposed to our participants show that in the egocentric condition (but not in the allocentric condition) subjects reported the feeling that the alien hand belonged to themselves and that they were the agent of the tapping action. Accordingly, we have found that cortico-spinal excitability for the own hand, as indicated by the MEPs value, was very similar in ego and solo condition. It is worth noting that we obtained these results without following the classical procedure to induce the RHI. In this respect other studies have shown this possibility (Kalckert and Ehrsson 2014). Similar results were found by Schutz-Bosbach et al. (2006). They used the RHI paradigm in order to determine if and how the motor system has the resources to distinguish between the self and the other's movements. Once the embodiment of the partner's hand was induced (in the synchronous condition), the MEPs facilitation, usually present during action observation paradigm and replicated in the no-embodiment (asynchronous) condition of their study, was abolished and no difference with respect to the baseline was found. Even more interesting, in our study, is the finding that MEPs increased in the allocentric condition, indicating an increase in the cortico-spinal excitability. In keeping with our predictions, this indicates that the proto-musical context of our experiment has the characteristics of a (proto-musical) joint action, that is, of a motor act that, through the social interaction with the partner, aims at reaching the required rhythmic alternation. The higher cortico-spinal excitability in the allocentric position is in accordance with Novembre et al. (2012)' study in which pianists, though in an exclusively acoustic condition, showed higher MEPs in left FDI, ADM and ECR when they performed the right part of a piano piece together with a hidden partner performing the left part, rather than performing it alone. Actually, Novembre et al. 2012 found the same facilitation pattern in a 'mute' condition, in which the pianist playing the melody could not hear his partner playing the bass-line. Then, in this case, we can exclude that eco-neurons (auditory mirror neurons which activate when an action is simply heard, as if that action is accomplished by the observer himself, Kohler et al. 2002) played any role. But, since both joint-conditions, ego and allo, included auditory feedback, and only the latter showed a different excitability pattern, this remark can be extended to our experimental setting: eco-neurons are neither a necessary nor a sufficient condition to elicit higher MEPs.

It is worth noting that, contrary to our results, Maeda et al. (2002) found that MEPs' amplitude for hand movements in allocentric condition (hand pointing toward the observer) were lower than MEPs recorded in egocentric condition (hand pointing away from the observer). However, in the Maeda et al.'s experiment both conditions were shown on a computer screen, that is, in a context which was even less ecological than ours and those discussed previously (Schutz-Bosbach et al. 2006, Della Gatta et al. 2016): an image on a screen versus a more (a real hand) or less (a fake hand) biological object. In our experimental setting the social affordances (Koblisch and Sebanz 2008, Gallese and Sinigaglia 2010) offered by a partner in an allocentric condition, the possibilities of enacting a joint (proto-musical) action, are quite richer than those offered by a partner in an egocentric position (least of all on a screen). This is possibly the reason why in the latter case MEPs turned out to be lower, while the sense of agency and ownership was higher, exactly as in the solo tapping condition.

Now, an interesting question that deserves further exploration is whether the phenomenon we are dealing with can be framed within the "minimal architecture for joint-action" (Vesper et al. 2010, Butterfill 2017). According to this model, a joint-action is made possible by three factors: representations, processes and "coordination smoothers", the first one being the goal of a joint-action (e.g. playing together), the second one being monitoring and predicting the unfolding and the outcome of such action (e.g. checking for rhythmic coherence of the ensemble), the third one being any behaviour facilitating the accomplishment of the action (e.g. slowing down one's own time, if it is perceived as faster than the other musicians' time). In our experiment the task wasn't explicitly social, in that participants were only told to synchronize to certain metronome beats (either the odd or the even beats), then a representation of the action as a joint-action wasn't explicitly required. Let's compare our task with Loehr & Vesper (2015) experiment in which a pair of non-musicians was instructed to learn either the melody or the accompaniment part of a simple piece of music. Once learned, each subject was asked to play its part either alone or with his/her complementary part. The authors take the higher rate of errors in the former condition (compared to the latter condition) as evidence that a co-representation of the joint-action was active, leading the subject to produce more mistakes in the alone condition. Nevertheless, both the behavioural and the physiological outcomes in our set-up suggest that an implicit shared motor representation emerged, insofar as mutual adaptive timing and high cortico-spinal excitability are well known markers of a social interaction. Moreover, such interaction can be conceived of as a special kind of bimanual action, in which "anticipation of another's action and preparation for your own are not two separate things […] in the same sense that, in preparing to perform a bimanual action, preparation for the actions to be performed by the left hand and anticipation of the movement of the right hand are parts of a single process" (Butterfill 2018, for some empirical evidence see Kourtis et al. 2013 and 2014). However, it is important to notice that in our egocentric condition (the most similar to a bimanual action), the cortico-spinal excitability was comparable to the solo condition, that is, to a condition that doesn't require coordination with any other agent (be it one's own left body part or someone else's), contrary to our allocentric condition.

Conclusion

In conclusion, our experiment showed that IE as MAT can be found also in non-expert musicians and it is still present when the spatial position of the partner's body affects the sense of body ownership and agency, thus indicating a universal value of such a form of proto-musical competence. Moreover, we showed that when the context induces an embodiment of the partner's hand, the subject's cortico-spinal excitability is similar to the solo condition. However, when the tapping is carried out with the partner in the allocentric condition, not only the body ownership and agency are not affected, but the subject's cortico-spinal excitability increases. This is a very interesting result because it shows that, while in a joint-tapping task the motor system distinguishes between the body self and the other's body, when a subject performs an action which is strongly related to the partner's action, a shared motor representation is activated in order to deal with the social context in which individuals co-act, possibly mediated by the mirror neuron system.

Acknowledgements

We are grateful to Ivan Schepers and Joren Six for their technical support and to Marta Omodeo Zorini for her help in the data collection.

References

Blanke, O., Slater, M., Serino, A. (2015). Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. *Neuron*, *88*(1), 145–166

Bolt N. & Loehr J. (2017). The predictability of a partner's actions modulates the sense of joint agency . *Cognition*, 161, 60-65

Botvinick M, Cohen J. (1998). Rubber hands ''feel'' touch that eyes see. *Nature*. 391:756

Bruno, V., Fossataro, C., Bolognini, N., Zigiotto, L., Vallar, G., Berti, A., & Garbarini, F. (2017). "The role of premotor and parietal cortex during monitoring of involuntary movement: a combined TMS and tDCS study." *Cortex*, 96, 83-94

Bruno, V., Fossataro, C., & Garbarini, F. (2017). Report of seizure induced by 10 Hz rTMS over M1. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2017.11.003

Bucchioni, G., Fossataro, C., Cavallo, A., Mouras, H., Neppi-Modona, M., Garbarini, F. (2016). Empathy or Ownership? Evidence from Corticospinal Excitability Modulation during Pain Observation. *Journal of Cognitive Neuroscience*, *28*(11), 1760–1771.

Butterfill S. (2018) Coordinating joint actions in Kirk Ludwig & Marija Jankovic (eds.), *Routledge Handbook on Collective Intentionality* Routledge

Clayton, Martin (2012) What is entrainment? Definition and applications in musical research, *Empirical Musicology Review.* 7 (1-2). 49-56

Colling L. & Williamson K. (2014). Entrainment and motor emulation approaches to joint action: Alternatives or complementary approaches? *Frontiers in Human Neuroscience*, Volume 8 (754)

Cross I. (2014), Music and communication in music psychology, *Psychology of Music*, V. 42: 6: 809-819

Decety, J., Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition.* 12, 577–596

Della Gatta F., Garbarini F., Puglisi G., Leonetti A., Berti A., Borroni P. (2016). Decreased motor cortex excitability mirrors own hand disembodiment during the rubber hand illusion. *eLife* 5:e14972

Doffman, M. R. (2008). *Feeling the groove: Shared time and its meanings for three jazz trios*. Ph.D. Dissertation, Music Department, Open University

Ehrsson HH, Spence C, Passingham RE. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305:875–877

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*. 73, 2608–2611

Fairhurst MT, Janata P, Keller PE. (2013). Being and feeling in sync with an adaptive virtual partner: brain mechanisms underlying dynamic cooperativity. *Cerebral Cortex* 23, 2592 – 600

Fossataro, C., Bucchioni, G., D'Agata, F., Bruno, V., Morese, R., Krystkowiak, P., & Garbarini, F. (2017). Anxiety-dependent modulation of motor responses to pain expectancy. *Social Cognitive and Affective Neuroscience*, 1–10

Fossataro, C., Bruno, V., Gindri, P., Pia, L., Berti, A., & Garbarini, F. (2017). Feeling touch on the own hand restores the capacity to visually discriminate it from someone else' hand: Pathological embodiment receding in brain-damage patients. *Cortex*. https://doi.org/10.1016/j.cortex.2017.06.004

Fossataro, C., Sambo, C. F., Garbarini, F., & Iannetti, G. D. (2016). Interpersonal interactions and empathy modulate perception of threat and defensive responses. *Scientific Reports*, *6*, 19353. https://doi.org/10.1038/srep19353

Freeman WJ. (2000) A neurobiological role of music in social bonding. In *The origins of music* (eds Wallin N.L., Merker B., Brown S.), 411 – 424. Cambridge, MA: The MIT Press

Gallese V, Sinigaglia C. (2010) The bodily self as power for action. *Neuropsychologia* 48:746–755

Garbarini, F., Fossataro, C., Berti, A., Gindri, P., Romano, D., Pia, L., … Neppi-Modona, M. (2015). When your arm becomes mine: pathological embodiment of alien limbs using tools modulates own body representation. *Neuropsychologia*, *70*, 402–13. https://doi.org/10.1016/j.neuropsychologia.2014.11.008

Garbarini, F., Fornia, L., Fossataro, C., Pia, L., Gindri, P., & Berti, A. (2014). Embodiment of others' hands elicits arousal responses similar to one's own hands. *Current Biology*, *24*(16), R738–R739. https://doi.org/10.1016/j.cub.2014.07.023

Garbarini, F., Piedimonte, A., Dotta, M., Pia, L., & Berti, A. (2013). Dissociations and similarities in motor intention and motor awareness: the case of anosognosia for hemiplegia and motor neglect. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(4), 416–419. https://doi.org/10.1136/jnnp-2012-302838

Goebl, W., Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception.* 26, 427–438

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 197–208.

Honing H., Cate C., Peretz I., Trehub SE. (2015) Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B* 370

Hove, M. J., Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27,* 949–961.

Kaplan, D. M., Bechtel, W. (2011). Dynamical Models: an alternative or complement to mechanistic explanations? *Top. Cogn. Sci.* 3, 438–444. doi:10.1111/j.1756-

8765.2011.01147.x

Keller, P. E., Knoblich, G., Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition.* 16, 102–111

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (205–221). Amsterdam: IOS Press.

Keller, P.E., Novembre, G., Hove, M.J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B, 369,* 20130394

Kelso J. A. (1995) *Dynamic Patterns*, MIT Press, Cambridge MA

Knoblich, G., Sebanz, N. (2008). Evolving intentions for social inter-action: From entrainment to joint action. *Philosophical Transactions of the Royal Society B, 363,* 2021–2031

Koelsch S., Gunter T., Friederici A., Schroeger E. (2000). Brain indices of music processing: “nonmusicians” are musical. *Journal of Cognitive Neuroscience*, Vol. 12/3, 520-541

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*(5582), 846-848

Konvalinka, I., Vuust, P., Roepstorff, A., Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *The Quarterly Journal of Experimental Psychology, 63,* 2220– 2230.

Kourtis, Dimitrios, Günther Knoblich, Mateusz Woźniak and Natalie Sebanz (2014). ‘Attention Allocation and Task Representation during Joint Action Planning’. *Journal of Cognitive Neuroscience* 26 (10): 2275–2286. doi:10.1162/jocn_ a_00634.

Kourtis, Dimitrios, Natalie Sebanz and Guenther Knoblich (2013). ‘Predictive representation of other people's actions in joint action planning: An EEG study'. *Social Neuroscience* 8 (1): 31–42. doi:10.1080/17470919.2012. 694823.

Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. S. (2008). Coordination dynamics of large-scale neural circuitry underlying rhythmic sensorimotor behavior. *J. Cogn. Neurosci.* 21, 2420–2433. doi: 10.1162/jocn.2008.21182

Leman M., Buhmann J., Van Dyck E. (2017) The empowering effects of being locked into the beat of the music in *Body, sound and space in music and beyond : multimodal explorations* Wöllner C. (ed.), *SEMPRE Studies in The Psychology of Music*. 13-28

Leman M. (2016), *The Expressive Moment*, MIT Press, Cambridge MA

Levitin D. (2006), *This Is Your Brain on Music*, Dutton Penguin, US-Canada

Loehr, Janeen D., and Cordula Vesper (2015). 'The sound of you and me: Novices represent shared goals in joint action'. *The Quarterly Journal of Experi- mental Psychology*, no. ja (June): 1–30. doi:10.1080/17470218.2015. 1061029.

Maeda, F., Kleiner-Fisman, G., Pascual-Leone, A. (2002). Motor facilitation while observing hand actions: Specificity of the effect and role of observer's orientation. *Journal of Neurophysiology*. 87, 1329–1335

Malloch S., Trevarthen C. (2009) *Communicative Musicality*. UK: Oxford University Press

Masumoto J., Inui N. (2014), Effects of speech on both complementary and synchronous strategies in joint action, *Experimental Brain Research*. Jul;232 (7):2421-9

McNeill WH. (1995) *Keeping Together in Time.* Cambridge, MA: Harvard University Press.

Merker, B. H., Madison, G., Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex, 45,* 4–17

Novembre G, Ticini LF, Schutz-Bosbach S, Keller PE. (2012) Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex* 22, 2894 – 2903

Novembre, G., Keller, P.E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience,* 8:603

Nowicki L, Prinz W, Grosjean M, Repp BH, Keller PE. (2013) Mutual adaptive timing in interpersonal action coordination. *Psychomusicology. Music Mind Brain* 23, 6 – 20

Overy, K., Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception, 26,* 489– 504

Pacherie E. (2012) The phenomenology of joint action: self-agency vs. joint-agency. In *Joint attention: new developments* (ed. A Seemann), 343 – 89. Cambridge, MA: MIT Press.

Phillips-Silver, J., Aktipis, C. A., Bryant, G. A. (2010). The ecology of entrainment: foundations of coordinated rhythmic movement. *Music Perception.* 28, 3–14.

Phillips-Silver J, Keller PE. (2012) Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*. 6, 26.

Pia, L., Garbarini, F., Fossataro, C., Burin, D., & Berti, A. (2016). Sensing the body, representing the body: Evidence from a neurologically based delusion of body ownership. *Cognitive Neuropsychology*.

Piedimonte, A., Garbarini, F., Rabuffetti, M., Pia, L., & Berti, A. (2013). Executed and Imagined Bimanual Movements: A Study Across Different Ages. *Developmental Psychology*, No Pagination Specified. https://doi.org/10.1037/a0034482

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology : 120*(12), 2008–39. https://doi.org/10.1016/j.clinph.2009.08.016

Repp BH, Su YH. (2013) Sensorimotor synchronization: a review of recent research (2006– 2012). *Psychonomic Bulletin and Review.* 20, 403 – 452

Schmidt RC, Richardson MJ. (2008) Dynamics of interpersonal coordination. In *Coordination: neural, behavioural and social dynamics* (eds A Fuchs, VK Jirsa), 281–308. Heidelberg, Germany: Springer.

Schoegler B. (2000), Studying temporal co-ordination in jazz duets, *Musicae Scientiae*, 75-91

Schutz-Bosbach S, Mancini B, Aglioti SM, Haggard P. (2006). Self and other in the human motor system. *Current Biology*. 16:1830--1834

Sebanz, N., Bekkering, H., Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10,* 70–76

van Noorden, L., Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research.* 28, 43–66

Vesper, C., Butterfill S., Knoblich G., Sebanz N. (2010). A minimal architecture for joint action. *Neural Networks* 23 (8-9): 998–1003. doi:10.1016/j.neunet.2010.06.002.

Wallin N, Merker B, Brown S. (2000). *The Origins of Music*. Cambridge, MA: MIT Press.

Zatorre, R. J., Chen, J. L., Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*, 547-558

Chapter 3. Does musical interaction in a jazz duet modulate peripersonal space?

Dell'Anna [1 2] A., Rosso [1 2,] M., Bruno [2] V., Garbarini [2] F., Leman [1] M., Berti [2] A.

Abstract

Peripersonal space, the space within reach, has been widely studied in the last twenty years, with a focus on its plasticity following the use of tools and, more recently, social interactions. Ensemble music is a sophisticated joint action that has been typically explored in its temporal rather than spatial dimension, even within embodied approaches. Therefore we devised a new paradigm in which two musicians could perform a jazz standard either in a cooperative (correct harmony) or in an uncooperative (incorrect harmony) condition, under the hypothesis that their peripersonal spaces would be modulated by the interaction. We exploited a well-established audio-tactile integration task as proxy for such a space. After the performances we measured reaction times to tactile stimuli on the subjects' right hand and auditory stimuli delivered at two different distances, (next to the subject and next to the partner). Since there is evidence that the integration of two different stimuli (e.g. a tactile and an auditory stimulus) is faster in the near, compared to the far space, in accordance with the relevant literature, we predicted that a cooperative interaction would have extended the peripersonal space of the musicians towards their partner, facilitating reaction times to bimodal stimuli not only in the near, but also in the far space. Surprisingly, we obtained the complementary result, that is, an increase of reaction times to the tactile-auditory near stimuli, but only after the uncooperative condition. This finding may be interpreted as a suppression of the subject's peripersonal space or as a withdrawal from the uncooperative partner. Subjective reports and correlations between these reports and reaction times are coherent with such interpretation. Finally, an overall better multisensory integration competence was found in musicians compared to a sample of non-musicians tested in the same task.

[1] IPEM, Ghent University, Belgium
[2] Department of Psychology, Turin University, Italy

Introduction

Music has the power to get people together. But how literally should we take this common sense statement? In the present study we focus on the possibility that peripersonal space can be modulated by a musical interaction, combining insights from both cognitive musicology and neuroscience. Peripersonal space, the space of interaction with the objects reachable by the hand, was defined for the first time by Rizzolatti et al. (1981) as the multisensory, body-part-centred representation of the space immediately surrounding the body. These authors discovered bimodal neurons endowed with overlapping visual and tactile receptive fields in the ventral premotor cortex, putamen and intra-parietal sulcus of monkeys: visual stimuli appearing, typically, close to the hand or the face, activate the same neurons activated by touching the hand or the face (see also Graziano et al. 1997). A growing amount of attention has been dedicated in the last twenty years to the study of peripersonal space, revealing also definitional and theoretical issues to accommodate contrasting empirical findings (Dijkerman & Farnè 2015, Hunley & Lourenco 2018, Bufacchi & Iannetti 2018). However, there is unanimous consensus about the plasticity of such a representation in so far as it has been shown to vary according to the use of tools (Iriki et al. 1996, Berti & Frassinetti 2000, Biggio et al. 2017, Bruno et al. 2019, see Brown & Goodale 2013 for a review) and, more recently, according to social interactions (Teneggi et al. 2013, Patanè et al. 2016, Pellencin et al. 2018). In what follows, we emphasize the latter of these aspects with the aim to explore the social impact of a jazz duo interaction on the space representation between two musicians.

Ensemble music is a sophisticated form of joint action that has been under intense investigation in the last ten years, since it provides a straightforward model of social interaction for both cognitive musicology and neuroscience (Keller et al. 2014, Walton et al. 2015, Eerola et al. 2018, Müller et al. 2018). While a small part of these studies has focused on big ensembles or orchestras (D'Ausilio et al. 2012, Badino et al. 2014), most of them tested dyads of musicians, taking advantage of the interesting trade-off between the ecological validity and the controlled set-up they offer (D'Ausilio et al. 2015). Dyads have been explored, among other things, to study temporal prediction abilities and motor simulation (Keller et al. 2007), synchronization of timing and motion (Goebl & Palmer 2009), coupling between neuronal oscillation and rhythm

(Lindenbergen et al. 2009), monitoring of joint actions and the relative neuronal markers (Loehr et al. 2013). As it is clear already by this short list, music and musical interactions have been studied mainly from a temporal viewpoint. Given the intrinsically temporal nature of music, aptly defined by Dissanayake (2000) as "the temporal art", the spatial dimension of musical interactions has so far been neglected (see also Schäfer et al. 2013). Nevertheless, the above-mentioned literature about the plasticity of peripersonal space, in particular after a social interaction, invokes an investigation of this issue also in the domain of ensemble music.

On the other hand, embodied approaches to music cognition (Leman 2007, Krueger 2014, Schiavio & De Jaegher 2017) have stressed the bodily components of interactions with music, which is one of the possible spatial dimensions of an interaction (Naveda & Leman 2010, Geeves & Sutton 2014). These approaches have also emphasized the mutual dependence of two or more interacting bodies during joint actions that require different degrees of complexity, ensemble music being one of the most complex. In order to capture this phenomenon, Fuchs & De Jaegher (2009) draw an analogy with tool use and put forward the concept of 'mutual incorporation': as tools can be incorporated into one's body-schema after a prolonged practice (Maravita & Iriki 2004), two interacting bodies may incorporate each other, adapting their behaviour to the partner's on the fly (Soliman et al. 2015). There have been several studies concerning mutual adaptation in musical contexts (Loehr & Palmer 2011, Zamm et al. 2014), but, again, with an emphasis on the temporal, rather than the spatial side. Finally, it is by now clear that interpersonal synchrony established by tapping or drumming increases the feeling of affiliation also in non-musicians, both children and adults (Kirschner & Tomasello 2009, Hove & Risen 2009, see also Stupacher et al. 2017). In particular, interpersonal entrainment, the spatiotemporal coordination between two or more individuals, often in response to a rhythmic signal (Phillips-Silver & Keller 2012, Clayton 2012), plays a central role in rituals and public happenings, promoting joint actions and social bonding (Kokal et al. 2011, Cross 2014). However, despite this wealth of studies on musical interaction, the function of peripersonal space in music arguably deserves more attention, since its potential modulation may drive or be driven by pro-social factors.

To this effect, we reviewed the relevant cognitive neuroscience literature, which highlights the relationship between multisensory integration and peripersonal space. There is evidence that, when a sound occurs close to the subject, it typically speeds up his/her reaction times to a tactile stimulus (Serino et al. 2007, Canzonieri 2012, see Spence et al. 1998 for visuo-tactile integrations in the near space). Since the facilitation occurs only in the near space, the same encoded by the above-mentioned bimodal neurons, it is fair to see it as a feature of the peripersonal space. Teneggi et al. (2013), exploiting a multisensory integration paradigm borrowed by Canzonieri et al. (2012), measured the space between two subjects by means of reaction times to a tactile stimulus (delivered, in this case, to the cheek) and to a dynamic sound, that is, a sound that gave the impression of moving either from the subject to the partner or the opposite way (manipulating its onset from two loudspeakers placed beside each person). In Teneggi et al (2013) the tactile stimulus was delivered at different temporal delays from the onset of the sound, thus occurring when the sound was perceived to be at different distances. In the first of their experiments, the authors showed that (vocal) reaction times to the audio-tactile stimuli increased when the subject faced a real person compared to a mannequin. This finding was interpreted as a shrinking of the subject's peripersonal space. In their second experiment, Teneggi et al. (2013) found that the effect of facilitation in the near space extended to the far space, that is, close to the partner (reaction times to tactile stimuli with sound close to the partner decreased), after a cooperative compared to an uncooperative interaction in an economic game. Moreover, in a third experiment, if one of the two loudspeakers was placed one meter behind the partner, who remained in the same position, after the cooperative condition the effect extended as to include the partner himself. The authors concluded that "high-level social and cognitive representations (e.g., cooperation) are immersed or recoded into the physical and perceptual experiences of the body, thereby providing concrete and rich feelings that facilitate prediction, evaluation, and social behaviour" (ibidem, p. 4).

In order to study whether a musical interaction in a dyad can modulate the musicians' peripersonal space, we took advantage of the multisensory integration paradigm put forward by Serino et al. (2007), a simplified version of Teneggi et al. (2013)'s, in that subjects had to respond (pressing a button) to a tactile stimulus delivered on the subject's right hand while a sound was played either near the subject or near his/her partner, without other intermediate sounds. This paradigm has been shown to induce a facilitation effect in the near space, whenever a tactile and an auditory stimulus close to the subject co-occur, contrary to when the sound come from the far space (the partner position) or when only a unimodal tactile stimulus is delivered. In order to compare a cooperative to an uncooperative condition, we devised a paradigm in which a musician was required to play a jazz standard (*Autumn Leaves*) with an accompanying guitarist (always one of the experimenters) who played either a correct or an incorrect chords sequence. Indeed, there is evidence that irregular chord functions presented in chord sequences are perceived as less pleasant than regular ones, eliciting amygdala responses related to the emotional value of sounds (Koelsch et al. 2008, Steinbeis et al. 2006). After both the cooperative and the uncooperative conditions, subjects performed the multisensory integration task with the partner in front of them and, as a baseline, they performed the same task with no partner in front of them (see Methods for details). Again inspired by Teneggi et al. (2013), we asked subjects about their partner's correctness, agreeableness, similarity to themselves, as well as the degree of dismay they felt playing with their partner, to confirm the efficacy of the manipulation. Moreover, given the possible impact of different levels of empathy on the perception of the space between the musicians, we administered the Davis' Interpersonal Reactivity Index.

The use of jazz music in our set-up was motivated by the manifold possibilities of interaction this genre allows, in so far as it heavily relies on improvisation (Walton et al. 2015, Iyer 2002). Thereby, after playing the main theme, the musicians were required to improvise on one or two chorus, with the recommendation not to stop until the end of the tune, not to verbally communicate and not to move from the place they were seated. We assumed that the possibility to improvise could have encouraged the collective composition of the tune in the cooperative condition, and, on the contrary, disrupted the fluency and the unity of the ensemble in the uncooperative condition. Recent neuroimaging studies on jazz musicians support the previous assumption since they found activation of default-mode regions of the brain and relative deactivation of executive control networks, which may allow the improviser to suspend conscious monitoring and enter a "flow-like" state (Limb & Braun 2008, but see Beatty 2015 for a review), favouring a collective musical outcome.

A side issue of the present study deals with the multisensory integration competences of musicians compared to non-musicians. We tested a sample of non musicians and compared their performance with our musicians' in a paradigm identical to the baseline condition of our experiment, that is, we compared the two groups in a reaction time task to a tactile stimulus, to an audio-tactile stimulus with sound close to the subject and to an audio-tactile stimulus with sound at 120 cm from the subject. Recently, Landry & Champoux (2017) showed that musicians derive a greater benefit than non-musicians from multisensory co-activation. The authors delivered tactile, auditory and auditory-tactile stimuli (with sounds always coming from the same distance, i.e. 60 cm) to a sample of non-musicians and to a sample of different musicians (for genres and instruments). Not only were the musicians faster in responding to the unimodal stimuli (both tactile and auditory), they were even better at rapidly integrating auditory and tactile stimuli. Indeed, the intensive training undergone by musicians induces well-known cortical-subcortical reorganizations that involve also sensory areas (Zimmermann & Lahav 2011, Kraus & White-Schwoch 2016). The fact that musicians are used to integrate sounds coming from different sources, for example in music ensembles, may impact on such ability.

To summarize, following Teneggi et al. (2013), our hypothesis predicted that a cooperative musical interaction could extend the musician's peripersonal space toward his/her partner, as measured by a facilitation effect not only in the near, but also in the far (extra-personal) space. On the other hand, we expected the baseline and the uncooperative condition to remain untouched by the manipulation, showing the typical facilitation effect in the near, but not in the far space. Finally, we expected musicians to exhibit faster reaction times than non-musicians to all the conditions of our multisensory task.

Materials and methods

Participants

Twenty-eight healthy participants took part in the experiment (17 males, mean age = 23,5, sd = 2,1). The sample size was selected after having performed an *a priori* power analysis in a pilot experiment of 10 non-musicians (6 females, mean age = 22, sd

= 2,5, see details below), performing the audio-tactile multisensory task during the baseline condition (see below, Experimental design and procedure). G*power software (www.psycho. uni-duesseldorf.de/abteilungen/aap/gpower3, Faul et al. 2007 and 2009) was used to estimate, in an *a priori* analysis, the sample size in a paired t-test (two-tailed), using the Wold's Cohens' d = 0.97; $\alpha = 0.05$; Power (1- β err prob.) = 0.90. A total sample size of 14 subjects was indicated (4 females, mean age = 25,2, sd = 4,1). Five of the musicians were singers, while the others were a trumpet player, three saxophone players and five guitarists. Furthermore, in order to compare the performance of the musicians (N=14) with a group of non-musicians, fourteen additional non-musicians were tested in our study, thus resulting in a new sample of fourteen non musicians (7 females, mean age = 22.3, sd = 2.3). All the participants were included in the experiment if they had no history of neurological, major medical or psychiatric disorders. With respect to musicians, only participants with at least 10 years of regular musical training, either formal or informal (mean = 13, sd = 5,9; years of expertise in ensemble music: mean = 9, sd = 5,9) and with the additional requirement to be able to play and improvise on the jazz standard *Autumn Leaves* were recruited. The experiment conformed to the principles of the declaration of Helsinki and informed written consent was obtained from each participant, who was naïve concerning the purpose of the study.

Experimental design and procedure

The musicians were advised to rehearse the jazz tune *Autumn Leaves* in E minor key before coming to the laboratory, pointing out that they were expected to play its theme and one or two chorus of improvisation on the harmonic structure, accompanied by a guitarist they did not know before (i.e., one of the experimenters, always the same). Each musician faced the following three conditions (Figure 1):

1) cooperative condition, in which the subject faced the partner at about 120 cm distance, playing the required tune twice at 120 bpm. The guitar accompaniment followed the right chord sequence and both players once they started were asked not to verbally communicate, not to move from their chairs and not to stop before the end of the tune. After each take, the participant performed the audio-tactile multisensory task (see details in *Experimental task*);
2) uncooperative condition, in which the subject played like in the cooperative

condition, except for the accompaniment. The partner, indeed, played a systematically altered chord structure, transposing it half a tone high and half a tone low every four correct bars. The participant performed the audio-tactile multisensory task (see details in *Experimental task*) after each of the two takes;

3) baseline condition, in which the subject seated alone facing the wall 120 cm in front of him/her and performed the audio-tactile multisensory task (see details in *Experimental task*).

Overall, we had two sessions, one with the cooperative and the other one with the uncooperative condition, whose order was counterbalanced. Each session took less than one hour each: about fifteen minutes to find the tactile threshold, ten minutes to perform twice, twenty minutes to accomplish the experimental task twice (after each performance), five minutes to fill in the questionnaire. After the first session, we asked the musicians to come back to the laboratory in two to three weeks for the second session, in order to avoid any learning effect in the experimental task. Condition 3), the baseline, was performed either at the beginning of the first session or at the end of the second session to avoid an order effect, taking about twenty minutes (Figure 2, left). After either condition 1) or 2), a questionnaire was administered containing Davis' Interpersonal Reactivity Index (IRI) and, more importantly for our purposes, after both conditions 1) and 2) the following questions concerning the partner were posed: 1) how correct did you find your partner's playing? 2) How much did you enjoy your partner's playing? 3) How similar to you did you find your partner? 4) How much dismay did you feel in playing with your partner? (on a scale 1 to 5). The group of non musicians only performed the baseline condition.

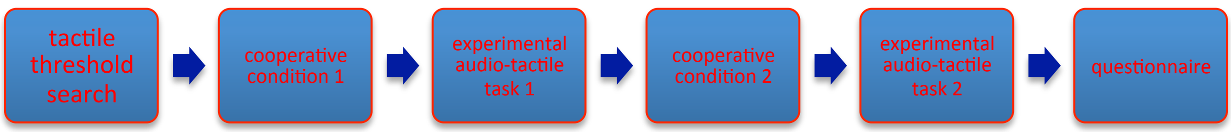


Figure 1. Timeline of the experiment. If the cooperative condition was performed on the first day, the uncooperative condition had to be performed on the second day and vice-versa. The baseline takes place either at the beginning of the first day or at the end of the second day. Non-musicians only performed the baseline condition.

Experimental task

Participants were comfortably seated on a chair. The individual tactile threshold to electrical non-painful somatosensory stimuli was found for each participant before starting the session. Beside the subject's chair, a response box lied on a desk, so that the subject's right index finger could push a button whenever a tactile stimulus was delivered on his/her right hand by means of bipolar skin electrodes (see details in *Stimuli*). Two loudspeakers were placed in the room: one just beside the subject's right arm, lying on the desk, one was placed 120 cm apart, beside a second chair (i.e., the partner's position). During the audio-tactile multisensory task, five different stimuli could occur (Figure 2, right): unimodal audio stimulus coming from the loudspeaker near to the participant (A Near); unimodal audio stimulus coming from the loudspeaker 120 cm far from the participant's chair (A Far); unimodal tactile stimulus delivered to the right hand dorsum of the participants' hand (T); bimodal audio-tactile stimuli consisting of a tactile stimulus delivered to the participants hand and a simultaneous audio stimulus coming from the loudspeaker near to the participant (TA Near); bimodal audio-tactile stimulus consisting of a tactile stimulus delivered to the participants hand and a simultaneous audio stimulus coming from the loudspeaker 120 cm far from to the participant (TA Far). The participants were instructed to respond only when they perceived a tactile stimulus, since we were interested in the facilitation effect of sounds on it. We used unimodal tactile stimuli to normalize (see *Data analysis*) the bimodal stimuli and used the unimodal auditory stimuli as catch trials. Reaction times (RT) and ratings (Likert scale from 0-7) about the intensity of the perceived tactile stimulus were collected. Ratings about the intensity were recorded to keep the subject's attention high during the whole task. The task consisted of two blocks of 60 trials each (i.e. 12 trials A Near, A Far, T, TA Near, TA Far), presented at a jittering inter-stimulus interval between 6000 and 8000 ms. The order of the trials' presentation was pseudo-randomized to avoid that more than two identical trials appeared in sequence. Stimulus presentation, synchronization and RT recording were controlled by E-prime v.2 (Psychology Software Tools, http://www.pstnet.com).

Stimuli

Somatosensory stimuli were transcutaneous electrical stimuli consisting in constant current square-wave pulses (DS7A, Digitimer) delivered to the right hand dorsum, using surface bipolar electrodes (1 cm between electrodes) attached to a Velcro strap. The stimulus duration was 200 μs. The stimulation intensity was adjusted and set at two-fold the individual perceptual threshold, estimated using the methods of limits (Gescheider, 1997), so that participants always perceived the tactile stimulation, which was never painful. The site of the stimulation was shifted by randomly displacing the electrodes' position of about 1 cm on the hand to prevent habituation. The mean stimulation intensity was 2.01 mA, range 3.2–0.51 mA. Auditory stimuli comprised 50 ms of tones of 784 Hz (Shrem et al., 2017), with an amplitude modulated sinusoidally at 50 Hz, including 5 ms rise and fall times. The stimuli were generated with Audacity software (http://audacityteam.org/) and presented via one or the other of the two loudspeakers.

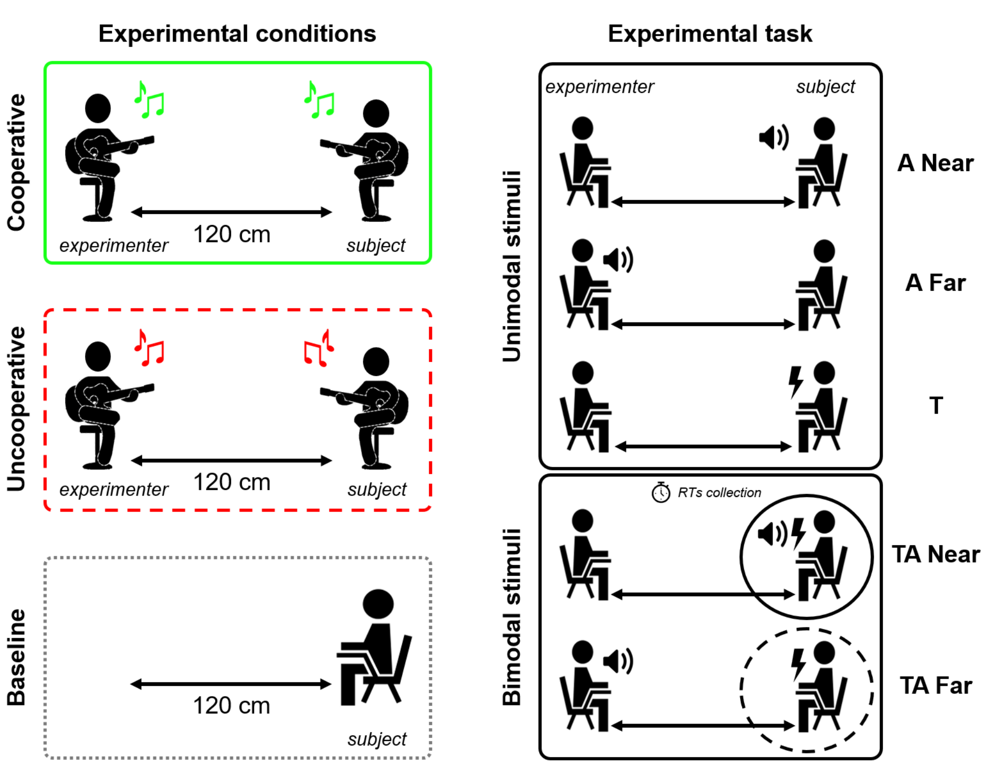


Figure 2. Experimental conditions and task. Musicians played in the cooperative (correct chord sequence played by the partner) and in the uncooperative (incorrect chord sequence) condition. After each of the two experimental conditions, musicians underwent the experimental task in front of their partner, pressing a button whenever a tactile stimulus was delivered on their right hand (with or without a sound). In the baseline condition, musicians (and, on different days, non-musicians) underwent the experimental task with no one in front of them.

Data analysis

Mean RTs for each condition of interest (i.e. T, TA Near, TA Far) were calculated for each participant in the main and in the control experiments. For each participant, in each condition, all trials in which the RTs were ±2 SD of the mean amplitude were identified as outliers and excluded from the analysis (in the Baseline condition: T (5.8%), TA Near (5.1%), TA Far (4.9%); in the Cooperative condition: T (5%), TA Near (5.1%), TA Far (4.6%); in the Uncooperative condition: T (5.6%), TA Near (5.4%), TA Far (6.2%). Since the rationale of the AET paradigm is to investigate the multisensory facilitation on the tactile detection induced by the concomitant presence of a sound inside the PPS (as in bimodal TA Near stimuli), we expressed data of the bimodal stimuli (i.e. TA Near and TA Far) as a ratio on the unimodal T stimuli (i.e. TA Near/T and TA Far/T) to compare RTs to different stimuli between the experimental sessions (i.e. Baseline, Cooperative, Uncooperative). Values lower than one represent a facilitation of the bimodal TA stimuli with respect to the unimodal T stimuli. These ratios were entered in a two-way repeated measure ANOVA with STIMULUS (TA Near/T; TA Far/T) and CONDITION (baseline; cooperative; uncooperative) as within-subject factors. Residuals were normally distributed according to Shapiro–Wilk test ($p > 0.05$) and the ANOVA ran properly. Post-hoc comparisons were carried out by means of Duncan's test (see *Supplementary Materials* for the analysis of the raw data).

We also performed t-tests on the four questions' scores concerning the partner, comparing them in the cooperative and uncooperative condition. Then, we carried out correlations between RTs and questionnaire' scores as follows: we computed an index of the cooperation effectiveness by subtracting values in the uncooperative condition from those in the cooperative condition in both RTs and subjective scores at the questionnaire. Note that for the RTs values we used normalized data (i.e. TA Near/T; TA Far/T; TA Near/TA Far). The size of these two indexes, then, depended on the difference between uncooperative and cooperative condition: the bigger the size, the higher this difference. Furthermore, we also explored correlations between the normalized RTs values and the values of the IRI questionnaire. Finally, we compared the raw RTs (not the ratios) of the group of non musicians with the raw RTs of the group of musicians in the Baseline condition. In order to control for the possible learning effect in the Baseline condition, we divided the group of musicians in two sub-groups, that is, the musicians who performed the Baseline at the beginning and the musicians who performed the Baseline at the end of the experiment. Hence, a mixed ANOVA was performed with STIMULUS (T vs TA Near vs TA Far) as within-subject factor and GROUP (musicians-baseline1 vs musicians-baseline2 vs non musicians) as between-subject factor.

Results

The two-way ANOVA (Figure 3) found a significant interaction effect between the type of STIMULUS and the type of CONDITION, $F(2,26) = 3.57$, $p = 0.04$, $\eta^2 = 0.20$. This indicates that the difference between RTs to TA Near and TA Far stimuli depends on the condition (baseline, cooperative, uncooperative). At Duncan test, in the cooperative as well as in the baseline condition TA Near was significantly lower than TA Far (cooperative: $p = 0.0024$/ baseline: $p = 0.04$), while TA Near and TA Far in the uncooperative condition were not significantly different (p=0.30). Furthermore, TA Near in the cooperative condition was significantly lower than TA Near (p=0.002) and TA Far ($p = 0.0003$) in the uncooperative as well as than TA Far (p=0.002) in the baseline condition. The remaining differences were non-significant. These results show that the effect of facilitation in the near space was present in the baseline and in the cooperative condition, but it was lost in the uncooperative condition (see *Supplementary Materials* for the results of the analysis of the raw data).

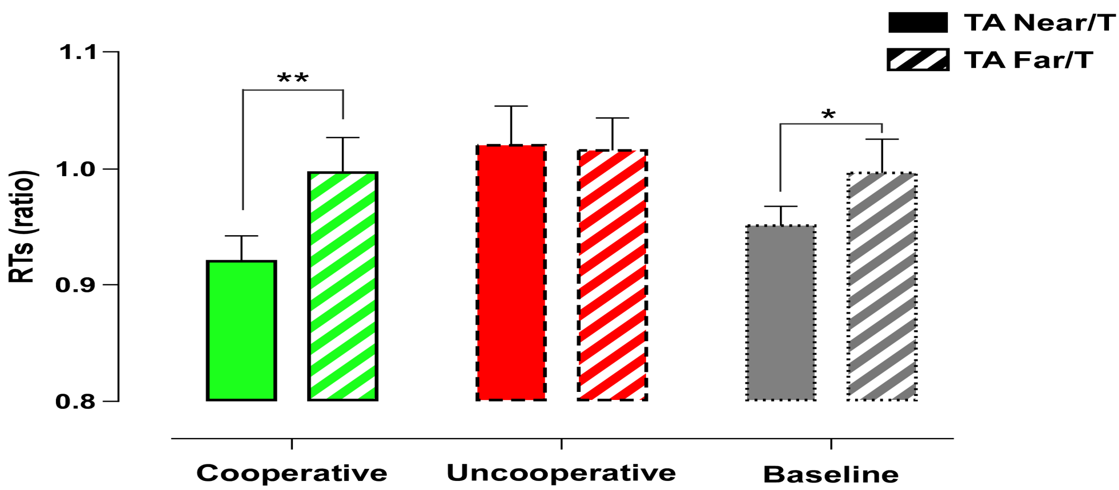


Figure 3. Response facilitation in the PPS. 2x3 ANOVA results (significant Interaction Stimulus by Condition). Mean ratios of RTs are significantly lower when the auditory stimulus is presented in the near space as compared to the far space in cooperative and baseline conditions. In the uncooperative condition, no significant difference was observed between near and far space. Error bars represent the standard error of the mean (*p<0.05, **p<0.01).

With respect to the questionnaire ratings (Figure 4A), when comparing the cooperative to the uncooperative condition, we found significant differences in the (two-tailed) t-tests in the answers to the first ("how correct did you find your partner's playing?"), $t_{(14)} = -3.29$, $p = 0.006$, second ("how much did you enjoy your partner's playing?), $t_{(14)} = -4.05$, $p = 0.001$ and fourth ("how much dismay did you feel in playing with your partner?), $t_{(14)} = 3.55$, $p = 0.004$, question of the questionnaire. This means that the experimental manipulation was effective from the subjective point of view, because the level of correctness and agreeableness was higher and the level of dismay lower in the cooperative compared to the uncooperative condition. When considering the correlation between the RTs and the questionnaire' scores (Figure 4B), we found only one significant correlation between the difference of the ratio TA Near/T in the cooperative and uncooperative condition and the difference of scores in answers to question 4) in the cooperative and uncooperative condition ($r = 0.58$, $p = 0.03$, $n = 14$). Even if this statistical result cannot survive to multiple comparisons, the present trend indicates that the bigger the difference in the near space facilitation between cooperative and uncooperative condition, the bigger the difference of dismay' score (question 4) between those conditions. Furthermore, a marginally significant correlation was found between the difference of the ratio TA Near/T in the cooperative and uncooperative conditions and the IRI values in the Perspective taking sub-scale ($r = 0.526$, $p = 0.05$, $n = 14$). Again, even if this statistical result cannot survive to multiple comparisons, the present trend indicates that the bigger was the difference between cooperative and uncooperative conditions (with greater advantage of TA Near with respect to TA Far in the cooperative than in the uncooperative condition) the higher was the level of empathy of the subjects. However, given the small sample, these correlations have to be considered as exploratory (see Table 1).

	Perspective taking	**Q1 UC-C**	**Q2 UC-C**	**Q3 UC-C**	**Q4 UC-C**
UC-C TA Near/T	***r* = 0.526** ***p* = 0.053**	*r* =-0.120 *p* = 0.68	*r* =-0.282 *p* = 0.32	*r* = -0.285 *p* = 0.32	***r* = 0.583** ***p* = 0.028**
UC-C TA Far/T	*r* =-0.039 *p* = 0.895	*r*= -0.289 *p* = 0.316	*r* =-0.391 *p* = 0.166	*r* = -0.275 *p* = 0.34	*r* = 0.517 *p* = 0.58
UC-C TA Near/TAFar	*r* = 0.053 *p* = 0.858	*r* = 0.179 *p* = 0.539	*r* = 0.117 *p* = 0.688	*r* = -0.017 *p* = 0.953	*r* = 0.089 *p* = 0.761
Perspective taking	//	*r* = 0.225 *p* = 0.439	*r* = 0.344 *p* = 0.229	*r* = 0.505 *p* = 0.66	*r* = -0.421 *p* = 0.134

Table 1. Correlations between RT indexes, questions about the partner and IRI (Perspective taking sub-scale). C=cooperative, UC=uncooperative.

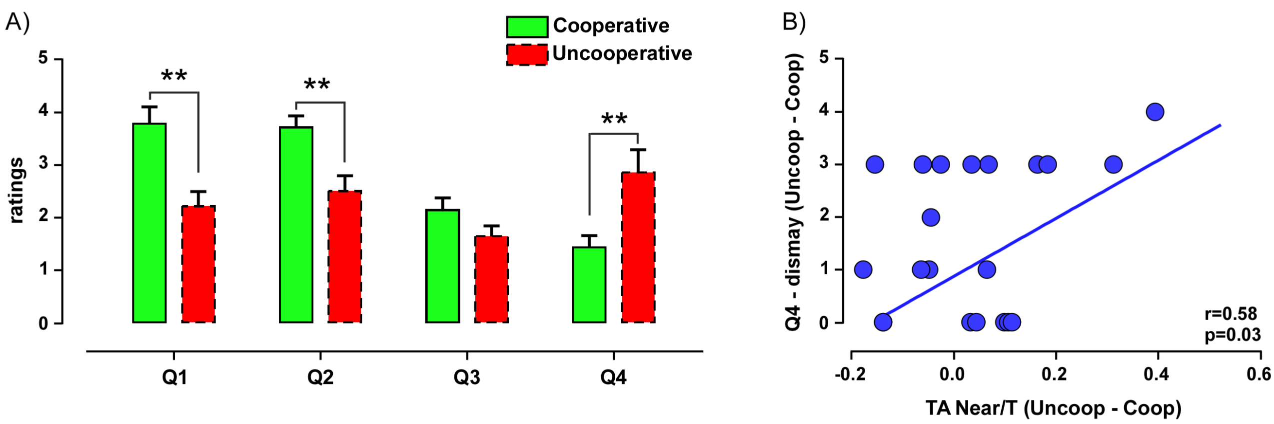


Figure 4. Questionnaire and correlation between RTs and subjective measures. A. *t-tests* differences between cooperative and uncooperative condition about levels of correctness (Q1), agreeableness (Q2) and level of dismay (Q4). Error bars represent the standard error of the mean (***p*<0.01). **B.** The graph shows a significant correlation between RTs in the near space and dismay caused by the partner.

Finally, with respect to the difference between musicians and non musicians

(Figure 5), our mixed ANOVA found a main effect of both GROUP, $F_{(2,25)} = 7.45$, p=0.003, η^2= 0.37 and STIMULUS, $F_{(2,50)} = 6.28$, $p = 0.003$, η^2= 0.2,. The Duncan's test shows that in all groups RTs to TA Near are significantly faster than RTs to T ($p = 0.0005$) and TA Far ($p = 0.0004$), but the effect of Group makes clear that musicians are always significantly faster than non musicians in both baseline1 ($p = 0.002$) and baseline2 ($p = 0.02$) sub-group. The interaction GROUP*STIMULUS was not significant ($F_{(4,50)} = 1.25$, p = 0.3, η^2= 0.09).

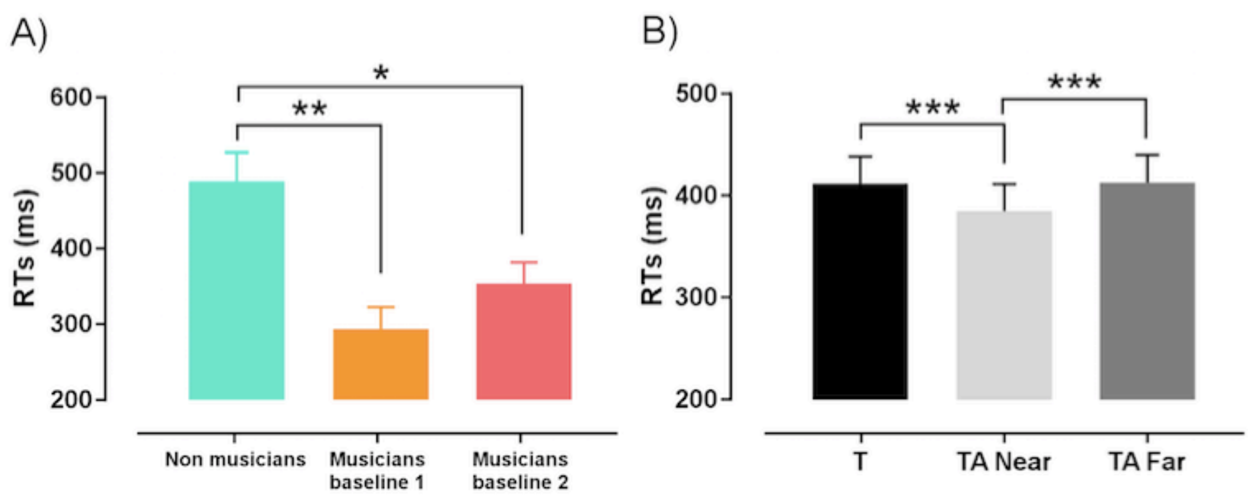


Figure 5. Comparison between musicians and non-musicians. 3x3 ANOVA results. A. Both sub-groups of musicians exhibit a significantly better performance overall with lower RTs, as addressed by the significant main effect of Group factor. **B.** Both groups showed a facilitation effect in the near space, i.e. significant main effect of Stimulus. Error bars represent the standard error of the mean (***p<0.001, **p<0.01, *p<0.05).

Discussion

In the present paper we investigated the topic of peripersonal space in a sample of musicians and, in particular, whether such a representation could be modulated by a duet jazz interaction, be it either cooperative or uncooperative. Interestingly, while our results did not show any modulation after the cooperative interaction (the typical facilitation effect in the near space, clearly visible in the baseline condition, did not extend to the far space), the complementary result emerged. In other words, after the uncooperative condition, reaction times in the near space lost the facilitation effect, becoming as slow as those in the far space. However unexpected, this result does not falsify our hypothesis. Basically, this result shows that the kind of interaction matters, although concerning the other side of the interaction, the uncooperative one. Crucially, the answers to the questions about the partner's correctness, agreeableness, similarity and degree of dismay indicated that the difference between cooperative and uncooperative conditions had a real subjective importance for the musicians, as confirmed also by the correlation between reaction times and answers to the question about the dismay caused by the partner's playing on the one hand, and (marginally) between reaction times and IRI values on the other hand. Finally, the comparison between musicians and non-musicians, showed an overall greater capacity of the former in the multisensory integration task we exploited.

The literature on peripersonal space has recently reported several remapping effects resulting from social interactions. Teneggi et al. (2013) showed that a cooperative economic game was capable to speed up the subject's response to audio-tactile stimuli in the far space. A similar pattern was found by Pellencin et al. (2018), according to which reaction times to visuo-tactile stimuli were reduced in the far space, if subjects faced a partner previously judged as moral (see also Iachini et al. 2015). Thus, a positive interaction seems to extend the peripersonal space up to include the other person in the subject's action area. The common sense idea that music connects people, supported by models of music considered as a "biotechnology of group-formation" (Freeman 2000), is coherent with anecdotic reports by musicians perceiving a "bubble" of sound around the ensemble, while they are playing together (Salice et al. 2017, see also Doffman 2008). All this led us to hypothesize that a cooperative musical interaction could have resulted in an enlargement of the musicians' peripersonal space toward the partner, arguably indicating that the members of the dyad are keen on musically interacting. A possible reason why this was not observed in our experiment may be related to the way of measuring peripersonal space. We used, indeed, Serino et al. (2007)'s paradigm, which consisted of only two points in space where the sounds were presented (the sound close to the subject and the sound close to the partner), instead of a series of points between the musicians (the dynamic sound), as in Canzonieri et al. (2012). Therefore, we cannot exclude that an effect might be observed using a more scaled sound presentation. Furthermore, at least two other peripersonal space (behavioural) measurements have been employed in previous studies, that is, line bisection with pencil or laser pointer (Berti & Frassinetti 2000, Bisio et al. 2017) and verbal reports of the distance at which an approaching partner was judged as reachable (Iachini et al. 2015, Patanè et al. 2016). The possibility of an extension of peripersonal space after a cooperative musical interaction by means of different measures, then, remains open (see also Bufacchi & Iannetti 2018 for similar remarks).

Our central finding, the loss of facilitation effect in the near space after the uncooperative condition, can be interpreted as a near space that becomes far, as if the subjects rejected the disturbing and upsetting partner or, rather, the subject himself "escaped away" from the partner. As a consequence, the partner would be no more suitable for (musically) interacting. It is worth to stress that, to our knowledge, this is the first finding interpretable as a disappearance of the space close to the subject. Pursuing the metaphor of the bubble, we might say that the multisensory bubble surrounding the musicians exploded after the uncooperative interaction. The concept of "mutual incorporation" may be useful to clarify our result from an embodied music cognition point of view (Leman 2007, Schiavio & De Jaegher 2017). When primates make prolonged use of tools, an embodiment of these tools into their body-schema occurs (Maravita & Iriki 2004). Similarly, musical instruments that are tools humans make use of for years, before they get mastered, may become incorporated into the musician's body-schema (see Nijs 2017 for an account of such a merging between the musician and his/her instrument, Clark 2008 for a thorough philosophical assessment of the extended mind hypothesis and Krueger 2014 for an application of such hypothesis to music perception). Inspired by Merlau-Ponty's philosophy, Fuchs and De Jaegher (2009) introduced the concept of "mutual incorporation" to draw an analogy between tool use and social interaction: as tools may be incorporated into the body-schema after their use, the interacting subjects may incorporate their partners to a degree dependent on the coordination between them, that is, from a simple weight lifting to playing or dancing together. Accordingly, Soliman et al. (2015) tried to experimentally demonstrate the existence of a joint body-schema emerging after a joint action (like sawing a candle with a string or rowing together). As we already pointed out in the Introduction, our work investigates the peripersonal space, rather than the body space or the body schema, and these three constructs need to be kept separated (see Hunley & Lourenco 2018). Still, a concept like "mutual incorporation" seems apt to emphasize the interdependence of both musicians' peripersonal spaces during and after the performance, well represented by the "bubble" metaphor. Such a metaphor may, indeed, accommodate Bufacchi & Iannetti (2018)'s warning against an "in-or-out" entity, that is a too sharp boundary between the peripersonal and the extra-personal space. These authors propose to substitute the bubble metaphor with the concept of gradient, since the latter seems able to do justice to the continuity of both neuronal and behavioural responses to the different multimodal stimuli tested in this research field. Nevertheless, nothing prevents from conceiving the bubble around one or more subjects as endowed with gradients that are plastically modifiable by both tools and social interactions: after all, the gradient itself has boundaries, that is, an in zone and an out zone.

The result we obtained in the uncooperative jazz interaction indirectly suggests that the dissonant chords played by the guitarist, repulsing the melodic line drawn by each soloist, caused an "explosion" of the bubble surrounding the two musicians, thus disrupting the mutual incorporation and the "group flow". Arguably, it might be also proposed that one of the components of the concept of flow along with the loss of time awareness (Csíkszentmihályi 1990) is the transformation of space perception and there have recently been attempts to apply the concept of flow to ensemble music (Hart & Di Blasi 2014, Cochrane 2017). Ethnographic studies reporting feeling of blending with the environment while listening to music support this view (Herbert 2011). Since sounds themselves have a spatial dimension, related to its propagation, also sounds musicians produce may contribute to shape the bubble around each of them and the ensemble. To this effect, our research questions would surely benefit from using 3D sounds in the future, since this technology allows to add a third dimension to the sound itself, emphasizing its spatial nature and making room for deeper soundscape explorations (see Brattico et al. 2017). For example, one may test how a condition in which music is heard through headphones (think of silent disco context) compared to a condition of global involvement in a 3D sounds environment (or to a condition like the one we studied) impacts on the feeling of connectedness of groups of musicians (or listeners).

A different interpretation might point to the fact that, while the harmony was altered, the remaining musical parameters (tempo, timing, timbre, structure of the tune, etc.) remained unaltered. Thus, an alternative is that the dissonance of the uncooperative condition led participants to ignore any audio stimuli not produced by themselves. If they kept ignoring stimuli not produced by them during the task, the lack of audio-tactile integration after the uncooperative condition also in the near space would be explained. However, analysis of the raw data that could speak to this possibility, like a higher similarity between uncooperative and baseline data compared to cooperative and baseline, turned out to be not significant, confirming the analysis done on the ratios (see *Supplementary Materials*). On the other hand, the marginally significant correlation between an index of the difference between uncooperative and cooperative condition in the reaction times to near stimuli and the IRI Perspective taking sub-scale values suggests that our manipulation was more effective the more empathic was the subject. Although this result has to be confirmed with a bigger sample than the one we tested, it highlights that factors other than music, e.g. personal traits, may affect peripersonal space modulation.

As we expected, subjective reports revealed a significant difference between cooperative and uncooperative conditions, with higher levels of correctness and agreeableness and lower level of dismay in the former compared to the latter condition. Moreover, a significant correlation was found between the level of dismay and reaction times to TA Near in the cooperative and uncooperative condition. Such a correlation indicates that the bigger is the difference between the two conditions in the responses to the bimodal stimulus in the near space, the higher is the level of dismay in the uncooperative compared to the cooperative condition. A small group of musicians turned out to appreciate the uncooperative more than the cooperative condition, and this might be explained by the fact that dissonant notes, although experienced as rejecting from a dyadic ensemble point of view, might nonetheless be intriguing and might trigger the musicians' higher inclination toward more musically experimental contexts (Torrance & Schumann 2018). Future studies might investigate this conjecture, comparing groups of musicians with different attitudes towards musical research. It might be speculated that musicians working with atonal music, electronics, free jazz are more fascinated than classical or pop musicians by dissonances and, as a consequence, might show an opposite pattern to the one we discovered, turning out to feel closer to a harmonically dissonant partner than to a consonant partner (Mencke et al. 2019).

Another avenue for further explorations has to do with the kind of instruments musicians use (Simoens & Tervaniemi 2013, Cheong & Will 2017). Arguably, being peripersonal space a body-part-centred representation, instruments that literally amplify a musician' spatial reaching (like drum sticks, keyboards, harps, etc.) need a separate treatment, not to mention singing, which does not employ any tool at all. We partly complied with the previous caveat, excluding from our experiment big size instruments like piano, drums, double bass, but we did include singers (see Kleber et al. 2016 for a study on the singer's brain). Nevertheless, a comparison between our singers and our instrumentalists samples did not show any difference. Moreover, since different body parts are involved according to the instrument played, different peripersonal space representations would likely ensue from the musician's hand, foot or mouth.

Finally, our finding that musicians react faster than non-musicians to both tactile and audio-tactile stimuli in both peripersonal and extra-personal space confirms recent results from Laundry & Champoux (2017). This is not surprising, given the intense multisensory training musicians undertake in learning to use their instrument and, to a less extent, their voice. Musicians need to reach a fine sensorimotor control to combine reading of a score (and, before learning it, looking at their actions toward the instrument), appropriately touching their instrument and listening to the sound they (or their partners) produce (Bangert et al. 2006, Jänke 2012, Reybrouck & Brattico 2015, for some neurophysiological evidence). However, it is worthwhile to note that in our experiment, as in Laundry & Champoux (2017), we did not test the visual modality, whose impact on the multisensory integration remains to be investigated. Moreover, contrary to Laundry & Champoux (2017), in our study the stimulated hand was the same that responded to the stimuli (the right hand), thus ruling out a possible role of the anterior corpus callosum, a typically more developed brain structure in musicians (Lee et al. 2003, Burunat et al. 2015).

To summarize, although our main hypothesis concerning a possible extension of a musician's peripersonal space after a cooperative jazz interaction with a partner was not entirely confirmed, a complementary and no less intriguing outcome emerged from the present study. The uncooperative interaction, in which the partner performed an altered guitar accompaniment, reduced the musician' speed in responding to the audio-tactile stimuli in the near space as if those stimuli came from the far space. In keeping with the recent literature about social modulation of peripersonal space, this result may be interpreted as a withdrawal from the upsetting partner from one' s space of (musical) interaction, suggesting that an embodied (jazz) performance has the capacity to remap the space between subjects, at least distancing them in an uncooperative condition.

We sincerely thank the musicians who enthusiastically took part in the experiment and Andrea Bussu for technical support.

References

Badino, L., D'Ausilio A., Glowinsky D., Camurri A. & Fadiga L. (2014) Sensorimotor communication in professional quartets. *Neuropsychologia* 55, 98–104

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, 30, 917–926.

Beaty, R. E. (2015). The neuroscience of musical improvisation. *Neurosci. Biobehav. Rev.* 51C, 108–117. doi: 10.1016/j.neubiorev.2015.01.004

Berti, A., & Frassinetti, F. (2000). When far becomes near: Remapping of space by tool use. *Journal of Cognitive Neuroscience*, 12, 415–420. https://doi.org/10.1162/089892900562237

Biggio M., Bisio A., Avanzino L., Ruggeri P. & Bove M. (2017) This racket is not mine: The influence of the tool-use on peripersonal space. *Neuropsychologia*, 103, 54-58

Bisio, A., Garbarini F., Biggio M., Fossataro C., Ruggeri P. & Bove M. (2017) Dynamic shaping of the defensive peripersonal space through predictive motor mechanisms: when the 'near' becomes 'far'. *J. Neurosci*. 37, 2415–2424

Brattico P., Brattico E. & Vuust P. (2017). Global Sensory Qualities and Aesthetic Experience in Music. *Front. Neurosci.* 11:159. doi: 10.3389/fnins.2017.00159

Brown LE & Goodale MA (2013) A brief review of the role of training in near-tool effects. *Frontiers in Psychology* **4**:576. doi: 10.3389/fpsyg.2013.00576

Bruno V., Carpinella I., Rabuffetti M., De Giuli L., Sinigaglia C., Garbarini F. & Ferrarin M. (2019) How tool-use shapes body metric representation: evidence from motor training with and without robotic assistance, *Frontiers in Human Neuroscience* (12)

Bufacchi RJ & Iannetti GD. (2018) An Action Field Theory of Peripersonal Space. *Trends in Cognitive Science*; 22:1076–1090. doi: 10.1016/j.tics.2018.09.004

Burunat I, Brattico E, Puoliväli T, Ristaniemi T, Sams M, Toiviainen P (2015) Action in Perception: Prominent Visuo-Motor Functional Symmetry in Musicians during Music Listening. *PLoS ONE* 10(9): e0138238.

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLoS ONE*, 7, e44306. https://doi.org/10.1371/journal.pone.0044306

Cheong Y.J. & Will U. (2018) Music, Space and Body: the evolutionary history of vocal and instrumental music. Conference: *ICMPC15-ESCOM10*

Clark A. (2008). *Supersizing the Mind*. Oxford University Press. Oxford

Clayton, M. (2012). What is entrainment? Definition and applications in musical research. *Empirical Musicology Review*. 7 (1–2), 49–56.

Cochrane, T. (2017). Group Flow. In Micheline Lesaffre, Pieter-Jan Maes & Marc Leman, ed. *The Routledge Companion to Embodied Music Interaction.* NY, USA: Routledge, pp. 133-140.

Cross, I., (2014). Music and communication in music psychology. *Psychology of Music* 42 (6), 809–819.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

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

D'Ausilio, A., Novembre, G., Fadiga, L. & Keller, P.E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, *19*(3), 111-114.

Dijkerman, H.C.& Farnè, A. (2015) Sensorimotor and social aspects of peripersonal space. *Neuropsychologia* 70, 309–312

Dissanayake E (2000). Antecedents of the temporal arts in early mother-infant interaction. In NL Wallin, B Merker, and S Brown, eds, *The Origins of Music,* pp. 389-410. MIT Press, Cambridge, MA.

Doffman, M.R., 2008. *Feeling the Groove: Shared Time and Its Meanings for Three Jazz Trios* (Ph.D. Dissertation). Ph.D. Dissertation, Music Department). Open University.

Eerola, T., Jakubowski, K., Moran, N., Keller, P. & Clayton, M. (2018). Shared Periodic Performer Movements Coordinate Interactions in Duo Improvisations. *Royal Society Open Science* 5(2): 171520.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Freeman, W.J., 2000. A neurobiological role of music in social bonding. In: Wallin, N.L., Merker, B., Brown, S. (Eds.), *The Origins of Music*. Cambridge, MA: The MIT Press, pp. 411–424.

Fuchs, T., & De Jaegher, H. (2009). Enactive Intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8*(4), 465-486

Geeves, A. & Sutton, J. (2014). Embodied cognition, perception, and performance in music. *Empirical Musicology Review, 9*(3-4), 247-253.

Gescheider, G.A. (1997). The classical psychophysical methods. In: *Psychophysics: The Fundamentals*, 3rd ed., pp. 45–72. Mahwah: Lawrence Erlbaum Associates.

Goebl, W. & Palmer, C., (2009). Synchronization of timing and motion among performing musicians. *Music Perception.* 26, 427–438.

Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77, 2268–2292.

Hart E. & Di Blasi Z. (2014) Combined flow in musical jam sessions: A pilot qualitative study. *Psychology of Music* 43(2):1-16 DOI: 10.1177/0305735613502374

Herbert R. (2011) *Everyday Music Listening: Absorption, Dissociation and Trancing*. Farnham: Ashgate

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27, 949–960.

Hunley SB. & Lourenco SF. (2018) What is peripersonal space? An examination of unresolved empirical issues and emerging findings. *Wiley Interdisciplinary Review Cogn. Sci.*;9:1–17. doi: 10.1002/wcs.1472

Iachini, T., Coello, Y., Frassinetti, F., & Ruggiero, G. (2014). Body space in social interactions: A comparison of reaching and comfort distance in immersive virtual reality. *PLoS One*, 9, e111511. https://doi.org/10.1371/journal.pone.0111511

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, 7, 2325–2330. https://doi.org/10.1097/00001756-199610020-00010

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception*, 19, 387–414.

Jäncke, L. (2012). The dynamic audio-motor system in pianists. *Annals of the New York*

Academy of Sciences, 1252, 246–252.

Keller, P.E., Knoblich, G., Repp, B.H., (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition.* 16, 102–111.

Keller PE, Novembre G, Hove MJ. (2014) Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Phil. Trans. R. Soc. B* 369: 20130394. http://dx.doi.org/10.1098/rstb.2013.0394

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102(3), 299–314.

Kleber B., Veit R., Moll C.V., Gaser C., Birbaumer N. & Lotze M. (2016). Voxel-based morphometry in opera singers: increased gray matter volume in right somatosensory and auditory cortices. *NeuroImage*, 133

Koelsch S., Fritz T. & Schlaug G. (2008) Amygdala activity can be modulated by unexpected chord functions during music listening, *NeuroReport*. 19(18): 1815-1819

Kokal I, Engel A, Kirschner S, Keysers C (2011). Synchronized Drumming Enhances Activity in the Caudate and Facilitates Prosocial Commitment - If the Rhythm Comes Easily. *PLoS One* 6(11): e27272. doi:10.1371/journal.pone.0027272

Kraus, N., & White-Schwoch, T. (2016). Neurobiology of Everyday Communication: What Have We Learned From Music? *Neuroscientist*. doi:10.1177/1073858416653593

Krueger, J. (2014a). Affordances and the musically extended mind. *Frontiers in Psychology*. *4*(1003). doi: 10.3389/fpsyg.2013.01003

Landry S.P. & Champoux F. (2017) Musicians react faster and are better multisensory integrators. *Brain and Cognition* 111, 156–162

Lee, D. J., Chen, Y., & Schlaug, G. (2003). Corpus callosum: Musician and gender effects. *NeuroReport*, 14, 205–209.

Leman, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press

Leman, M. (2012). Musical entrainment subsumes bodily gestures – its definition needs a spatiotemporal dimension. *Empirical Musicology Review, 7*, 63-67.

Limb, C. J. & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: an FMRI study of jazz improvisation. *PLoS ONE* 3:e1679. doi: 10.1371/journal.pone.0001679

Lindenberger U, Li SC, Gruber W, Müller V (2009) Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience* 10:22.

Loehr, J. D., Kourtis D., Vesper C., Sebanz N. & Günther K. (2013). "Monitoring Individual and Joint Action Outcomes in Duet Music Performance." *Journal of cognitive neuroscience* 25(7):1049–61.

Loehr, J. D. & Palmer C. (2011). "Temporal Coordination between Performing Musicians." *Quarterly Journal of Experimental Psychology.* 64 (11): 2153-67

Maravita, A. & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, 8, 79–86. https://doi.org/10.1016/j.tics.2003.12.008

Mencke I., Omigie D., Wald-Fuhrmann M. & Brattico E. (2019). Atonal Music: Can Uncertainty Lead to Pleasure? *Front. Neurosci*. 12:979. doi: 10.3389/fnins.2018.00979

Müller V., Delius JAM., Lindenberger U. (2018). Complex networks emerging during choir singing. *Annals of the New York Academy of Sciences*. 1431(1):85-101.

Naveda, L. & M. Leman (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). *Music Perception 28*(1), 93–111.

Nijs, L. (2017). The merging of musician and musical instrument: an internal model-based approach. In M. Lesaffre, M. Leman & P. J. Maes (Eds.), *Routledge Companion to Embodied Musical Interaction*. London: Routledge.

Patané, I., Iachini, T., Farnè, A., & Frassinetti, F. (2016). Disentangling action from social space: Tool-use differently shapes the space around us. *PLoS One*, 11(5), e0154247. https://doi.org/10.1371/journal.pone.0154247

Pellencin, E., Paladino, M. P., Herbelin, B., & Serino, A. (2018). Social perception of others shapes one's own multisensory peripersonal space. *Cortex,* 104, 163–179.

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, 6, 26.

Reybrouck M. & Brattico E. (2015). Neuroplasticity beyond Sounds: Neural Adaptations Following Long-Term Musical Aesthetic Experiences. *Brain Sciences, 5*(1), 69-91.

Rizzolatti, G., Scandolara, C., Matelli, M. & Gentilucci, M., (1981). Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behav. Brain Res.* 2 (2), 147–163. https://doi.org/10.1016/0166-4328(81)90053-X

Salice A., Høffding S. & Gallagher S. (2017). "Putting Plural Self-Awareness Into Practice. The Phenomenology of Expert Musicianship." *Topoi* DOI 10.1007/s11245-017-9451-2

Schäfer, T., Fachner, J. & Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Frontiers in Psychology, 4,* 508.

Schiavio & De Jaegher (2017) Participatory sense-making in Joint musical practice. In *Routledge Companion to Embodied Music Interaction* (eds. Lesaffre M., Maes P.J., Leman M.), 31-39. New York and London: Routledge

Serino A., Bassolino, M., Farnè, A., & Làdavas, E. (2007). Extended multisensory space in blind cane users. *Psychological Science*, *18*(7), 642–648. https://doi.org/10.1111/j.1467-9280.2007.01952.x

Shrem T., Murray M. & Deouell L. (2017). Auditory-visual integration modulates location-specific repetition suppression of auditory responses. *Psychophysiology,* vol. 54, 11, 1663-1675

Simoens V. L. & Tervaniemi M. (2013). Musician–instrument relationship as a candidate index for professional well-being in musicians. *Psychology of Aesthetics, Creativity and the Arts*. 7(2), 171-180

Soliman, T. M., Ferguson, R., Dexheimer, M. S. & Glenberg, A. M. (2015) Consequences of joint action: Entanglement with your partner. *Journal of Experimental Psychology: General* 144(4): 873-888.

Spence, C., Pavani, F., & Driver, J. (1998). What crossing the hands can reveal about crossmodal links in spatial attention. *Psychonomic Societ*, 3(13).

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380

Stupacher J., Maes PJ., Witte M. & Wood G. (2017) Music strengthens prosocial effects of interpersonal synchronization – If you move in time with the beat, *Journal of Experimental Social Psychology* 72, 39–44

Teneggi, C., Canzoneri, E., di Pellegrino, G. & Serino, A. (2013). Social modulation of peripersonal space boundaries. *Current Biology*, 23(5), 406–411. https://doi.org/10.1016/j.cub.2013.01.043

Torrance, S & Schumann, F 2018, 'The spur of the moment: What jazz improvisation tells cognitive science' *AI and Society*, pp. 1-18. DOI: 10.1007/s00146-018-0838-4

Walton, A., Richardson, M. J., Langland-Hassan, P. & Chemero, A. (2015). Improvisation and the self- organization of multiple musical bodies. *Frontiers in Psychology*. 6, 313.

Zamm, A., P.Q. Pfordresher & C. Palmer. (2015). Temporal coordination in joint music performance: effects of endogenous rhythms and auditory feedback. *Exp. Brain Res.* 233: 607–615.

Zimmerman, E., & Lahav, A. (2012). The multisensory brain and its ability to learn music. *Annals of the New York Academy of Sciences*, 1252, 179–184.

Chapter 4. Touching the music. Integrating touch and sound of a trumpet. A TMS study

Dell'Anna A. [3] [4], Sarasso P. [4], Serra H. [4], Ricci R. [4], Berti A. [4]

Abstract

A sample of wind musicians and non-musicians was tested, under the hypothesis that higher FDI cortico-spinal excitability of the former group after single-pulse TMS could reveal their expertise in integrating (tactile) holding of a trumpet with the sound of a trumpet. Given the small sample, our results have to be considered as temporary, showing a trend towards a lower cortico-spinal excitability in musicians, compared to non-musicians, while listening to white noise, compared to trumpet tones, independently of the tool held. The lower sensitivity to white noise exhibited by musicians may be taken as a marker of their sensorimotor and ensemble music expertise, but further testing is needed to draw stronger conclusions about both populations' multisensory competences.

Introduction

"If humans are natural-born cyborgs (Clark 2003), tool use shape the way they know, that is, the way they are", might say a bold philosophical claim linked to the so called "extended mind hypothesis" (see also Clark & Chalmers 1998, Clark 2008 and Menary 2010 for an overview). Whatever the merits and limits of such hypothesis, a large amount of cognitive neuroscience studies in the last twenty years has highlighted how tool use modulates a wide range of phenomena like peripersonal space perception (Iriki et al. 1996, Berti & Frassinetti 2000, Serino et al. 2007, Brown & Goodale 2013, Patanè et al. 2016), body-schema (Maravita & Iriki 2004, Hunley & Lourenco 2018) and, more recently, multisensory integration (Lappe et al. 2008, Yamaguchi et al. 2013, Laundry & Champoux 2017). In the present article we investigated whether and how

[3] IPEM, Ghent University, Belgium
[4] Department of Psychology, Turin University, Italy
We sincerely thank the musicians who took part in the experiment, in particular the inimitable BandaKadabra.

musical tools, trumpets in particular, can modulate cortico-spinal excitability in wind musicians compared to non musicians, if a trumpet is held in the right hand while a trumpet sound, compared to white noise, is heard. This is not a familiar condition for non-musicians, but it is quite often experienced by wind musicians whenever they listen to someone else playing in their same ensemble or when they listen to a recording while practicing, be it for imitation or for correction of their own playing.

On the other hand, it is by now well known that musical training engages sensory and motor brain networks, inducing cortico-muscular changes (Munte et al. 2002, Bangert et al. 2006, Gentner et al. 2010, Herholz & Zatorre 2012, Schlaug 2915, Reybrouck et al. 2019) to such an extent that even only listening to practised music can increase activity in the fingers primary motor cortex of pianists (Haueisen & Knösche 2001, Jäncke 2012). Somatosensory cortical representations of fingers of the left hand are increased in string musicians, who use them a lot more than right hand fingers in playing violins or double basses, at least in classical music (Elbert et al. 1995), while harpists exhibit higher cortico-spinal excitability than non musicians for left FDI (first dorsal interossus), but not for ADM (abductor digiti minimi), which is not involved in plucking the harp's strings (Biuck et al. 2016). Playing wind instruments, in turn, recruits not only upper limbs, but also facial muscles like tongue, lips (orbicularis oris) and digastric muscle (Gotouda et al. 2007, Potter et al. 2015). Indeed, Schulz et al. (2003) found that bimodal tactile-auditory stimuli induced faster MEG responses in trumpet players somatosensory cortex compared to non musicians', but only when the tactile stimulus occurred on the lips, compared to the index finger (the co-occurring auditory stimulus was a trumpet tone). Multimodal integration may be seen as a peculiar trait of musical training in so far as it requires fast handling of auditory (sounds), tactile (instrument), visual (score and, in the initial learning phase, fingering) and proprioceptive stimuli in coordination with efferent signals. Accordingly, musicians have been labelled "better multisensory integrators" since they were shown to react faster than non musicians to bimodal tactile-auditory stimuli in a simple reaction times task in which tactile stimuli were delivered on the hand while co-occurring sounds were heard (Laundry & Champoux 2017, see also Dell'Anna et al. submitted for a similar result with sounds coming from the near, besides the subject, and the far, 120 cm apart, space).

In a rather different context, Yamaguchi et al. (2014) found higher cortico-spinal excitability in the masseter muscle when subjects saw images of food (compared to non-food images) while holding chopsticks (compared to scissors) in their hand, showing that congruent visual and somatosensory eating stimuli enhanced the neuromuscular system activation involved in eating actions. Interestingly, neither simply touching the chopsticks nor simply looking at the food images induced such excitability, suggesting that only the integration of several congruent sensory modalities is able to recruit cortical motor networks, at least in this experiment. This finding is in accordance with a theory of 'affordances' that takes into account not only the visual properties of objects (like Gibson's 1979), but also their auditory and somatosensory properties on the one hand, and the 'effectivities' of the subject, on the other hand. While affordances are the possibilities for action offered by an object to a subject (a tree to be climbed, a branch to be grasped and so on), effectivities are the organism's features allowing for interaction with objects, that is, muscular structure, movement capacities, size, shape, needs and sensitivities (Shaw & Turvey 1981). The interplay between affordances and effectivities has been recently explored by a couple of studies (Ranganathan et al. 2011, Sartori et al. 2011).

Now, consider musical tools. "A musical instrument is a type of transducer, converting patterns of body movement into patterns of sound . . . the interaction between the human body, with its intrinsic modes of operation, and the morphology of the instrument may shape the structure of the music" (Baily 1992, p. 149, see also Leman 2007). For example, the musician's fingers are suited for pressing the piston valves of the trumpet and the valves are built to afford such an action. Moreover, the sound produced by the trumpet may offer several action possibilities, like synchronization and movement (which is common reaction to every kind of music), but also emotional reactions that depend on its bright timbre and high pitch range (Windsor & de Bezenac 2012). Of course, some of these actions are dependent also on the expertise of the listener and may lead to shape his/her following tones according to the instrument played. If this instrument is precisely the trumpet, the whole neuromuscular system necessary to play it enter the scene, modulating lips, tongue, arms, hands and fingers muscles in order to obtain a given auditory outcome that suits the heard tones.

Coherently with the above arguments, we devised an experimental paradigm in which wind instruments musicians are compared to a sample of non musicians in a multisensory integration task. Both groups subjects held in their right hand either a trumpet or a scissors and listened either to one of five different pitched trumpet tones or to white noise, while single-pulse TMS was delivered on FDI left motor cortex. We hypothesized that, if wind instruments musicians are expert integrators of sounds and tools, sounds and tools being a trumpet in our case, higher cortico-spinal excitability would ensue for musicians compared to non musicians from the congruent condition, that is, when touching the trumpet while listening to trumpet tones.

Materials and methods

Participants

Nineteen right-handed volunteers (9 wind musicians, all males, 3 trumpet players, 3 saxophone players, 3 trombone players, mean age = 30.1 years, standard deviation = 6 and 10 non musicians, 5 females, mean age = 24.9, standard deviation = 1.9) took part in the experiment. A musician and a non-musician were excluded having too high a resting motor threshold (rMT). Participants were screened to exclude neurological or medical disease. The participants were naıve with regard to the purpose of the study. The experiment was approved by the Ethics Committee of the University of Turin and informed written consent was obtained from each participant.

Magnetic Stimulation

Motor evoked potentials (MEPs) were elicited by single-pulse TMS (Magstim Rapid2; Magstim Co. Ltd, Whitland, UK) with a 70-mm figure-of-eight-shaped coil positioned over the hand area of the right M1. The optimal location for stimulus induction (the location that gave the maximum MEP amplitude) for the right FDI muscle was identified. At this location, the coil was positioned and fixed with the handle pointing backwards at 45 degrees from the midline so as to activate the selected muscle. Then, the rMT was determined as the intensity needed to evoke a MEP in the relaxed muscle of more than 50 μV in 5 out of 10 consecutive trials. The stimulator output was set at 115% of each subject’s rMT (M non-musicians = 64 %, SD = 4.4, M musicians =

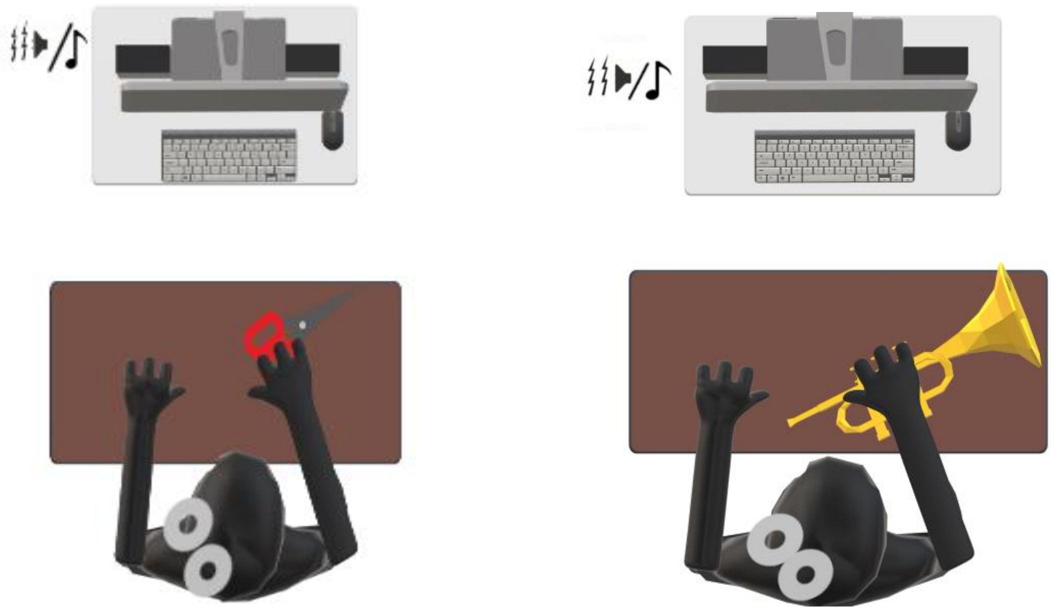
57.4 %, SD = 3, range 57-68 % of the maximum stimulator output). Participants who showed a rMT higher than 70% of the stimulator output were excluded from the stimulation phase.

Electromyography recording

Electromyographic (EMG) activity was recorded (MP150, Biopac System, USA), from the right first dorsal interosseous muscle (FDI) by self-adhesive bipolar surface electrodes with active electrode over the muscle belly and the reference electrode over the associated joint or tendon. Signals were amplified and digitalized with a sample rate of 10 kHz, filtered with a band-pass (10 Hz to 500 Hz), and stored in a computer for offline analysis, according to methods used in previous studies (Dell'Anna et al. 2018, Fossataro et al. 2018)

Task and procedure

Subjects were comfortably seated in front of a computer screen holding in their right hand either a trumpet or a scissors placed on a board on which also the left hand leaned, except for the baseline condition, in which they did not hold anything. The experiment was programmed using E-prime software V2.0 (Psychology Software Tool Inc., USA) in order to trigger TMS pulses at a controlled timing and trigger the EMG recording. In order to check for any cortico-spinal excitability change related to TMS per se, 10 baseline MEPs were recorded before (i.e. baseline-pre) and 10 after (i.e. baseline-post) the experimental block, with an inter-stimulus interval of 10 s. The MEPs amplitude recorded during the baseline was used to normalize data recorded during the experimental conditions. The E-prime script presented 30 randomized 50 ms sounds (15 trumpet tones, 3 for each of the following pitches, C, D, E, F, G and 15 white noise samples, built with Audacity Software) with an inter-onset interval of 10 s. After each sound a TMS pulse was delivered with a 250 ms delay from the sound onset. We ran the script four times (2 blocks while participants held the trumpet and 2 blocks while they held the scissors in their right hand, in a randomized order) resulting in 120 MEPs per subject, 30 for each of the following experimental conditions (Figure 1):



A B

Figure 1. Experimental conditions. Participants were tested by means of single-pulse TMS on FDI M1 while listening either to one of five differently pitched trumpet tones or to a white noise sample and holding either a scissor (A) or a trumpet (B).

1) Trumpet/Trumpet: each participant held a trumpet and listened to one of five differently pitched trumpet tones;
2) Trumpet/White noise: each participant held a trumpet and listened to a white noise sample;
3) Scissors/Trumpet: each participant held a scissors and listened to one of five differently pitched trumpet tones;
4) Scissors/White noise: each participant held a scissors and listened to a white noise sample.

For the purpose of keeping subjects attention high, they were asked to verbally evaluate the beauty of each tone on a Lickert-scale from 1 to 9 as soon as a cross appeared on the screen, 3 seconds after each sound onset. Once the subject's rMT was established, the experiment took approximately 40 minutes.

Data Analysis

EMG data were analysed offline using AcqKnowlege software (Biopac Systems, Inc., Santa Barbara, CA) to measure the peak-to-peak amplitude (in µV) and MEPs with an amplitude lower than 50 µV were discarded from analysis. Trials showing pre-activity (EMG signal greater than 50 µV) in the time window of 100 ms before the TMS pulse were excluded from analysis. Normal distribution of the residuals was checked using the Shapiro-Wilk test ($p>0.05$) and the appropriate parametric tests were performed SPSS Software 25 (IBM SPSS Statistics, USA). The mean MEPs values acquired during baseline-pre were compared to baseline-post by means of a paired t-test. According to the negative results in the baseline analysis, the mean MEPs amplitude of the baseline was used to normalize data of the experimental conditions. For the experimental condition a MEPs ratio (MEP ratio= $MEP_{obtained}/MEP_{baseline}$) was calculated and used as dependent variable in a mixed 2x2x2 ANOVA with "touch" (trumpet and scissors) and "sound" (trumpet and white noise) as within-subjects factors and "group" (musicians and non-musicians) as between-subject factor. Post hoc comparisons were carried out in case of significant results (Dell'Anna et al. 2018, Fossataro et al. 2018).

Results

Firstly, we found a significant difference between the two groups resting motor thresholds [$t_{(17)} = 3.67$, $p = 0.002$, two-tailed]. Specifically, the motor threshold was lower for the group of musicians (mean=57, standard deviation=3) than for the group of non-musicians (mean=64, standard deviation=4). In the baseline analysis, the t-test (two-tailed) comparing MEPs acquired during baseline-pre with those acquired during baseline-post did not show any significant difference ($t_{(37)}= 1.160$; $p = 0.253$). This indicates that TMS per se did not induce changes in cortico-spinal excitability, which remained unchanged from the beginning compared to the end of the experimental block. Since raw data were not normally distributed, we performed a log10 transformation (Osborne 2002) and then expressed the results as ratios (MEP ratio= $MEP_{obtained}/MEP_{baseline}$). Two and three outliers were then removed from the non-musicians and the musicians group respectively, resulting in a sample of 8 non-musicians and 6 musicians.

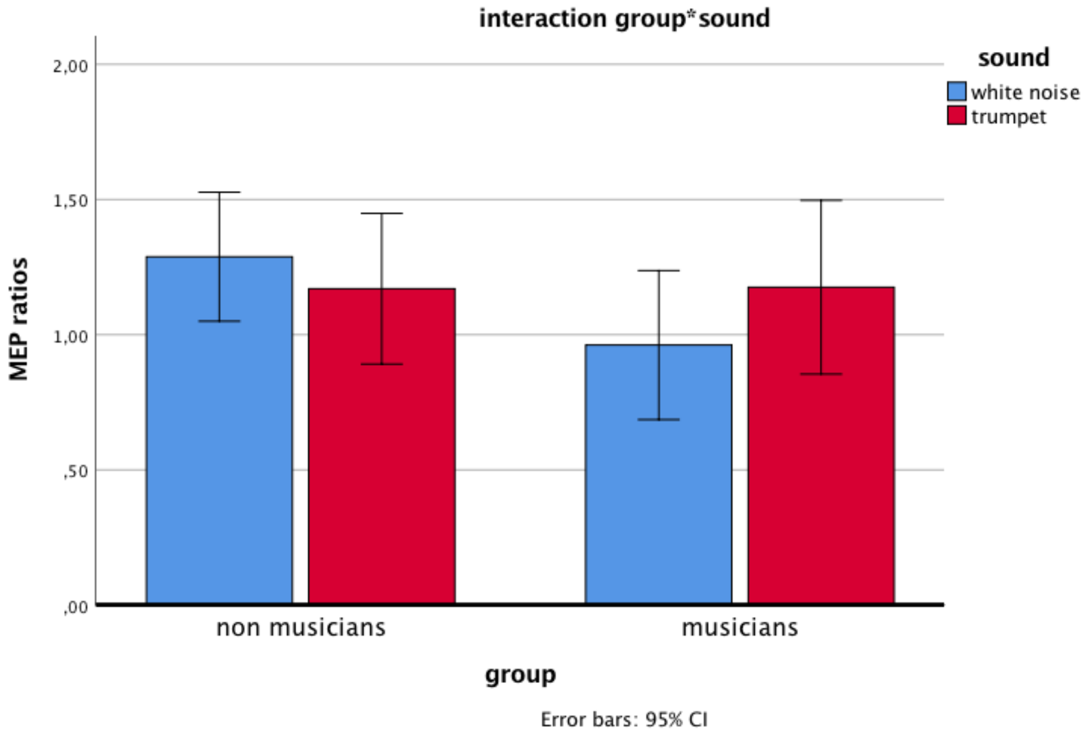


Figure 2. Interaction group*sound. Despite the significant interaction, no significant differences are revealed by post-hoc comparisons, even if a trend can be noticed toward higher cortico-spinal excitability in non-musicians listening to white noise, compared to musicians, independently of the tool held.

The mixed ANOVA showed a significant interaction between group and sound [$F_{(1,12)} = 4.77$, $p = 0.049$, $\eta^2 = 0.28$, power= 0.52], with a group showing different MEPs' modulation for a sound, independently of the tool held (Figure 2). However, post-hoc analysis did not show significant differences between conditions, although musicians seems less activated than non musicians by white noise (t-test, $t_{(26)} = 1.84$, $p = 0.077$, two-tailed). Also a significant interaction touch*sound was found [$F_{(1,12)} = 6.17$; $p = 0.029$; $\eta^2 = 0.34$; power= 0.62], indicating a different MEPs modulation according to the tool/sound combination, independently of the group (Figure 3). Nevertheless, post-hoc analysis did not show significant differences between conditions, even if a tendency could be noticed between higher MEPs listening to the trumpet rather than white noise, while holding a scissors (t-test, $t_{(14)} = -1.74$, $p = 0.1$, two-tailed). All the remaining results were not significant, since there was no main effect of either touch [$F(1,12) = 0.28$, $p = 0.6$] or sound [$F(1,12) = 0.39$, $p = 0.54$] or interactions touch*group [$F(1,12) = 2.86$, $p = 0.11$] or group*touch*sound [$F(1,12) = 1.37$, $p = 0.26$].

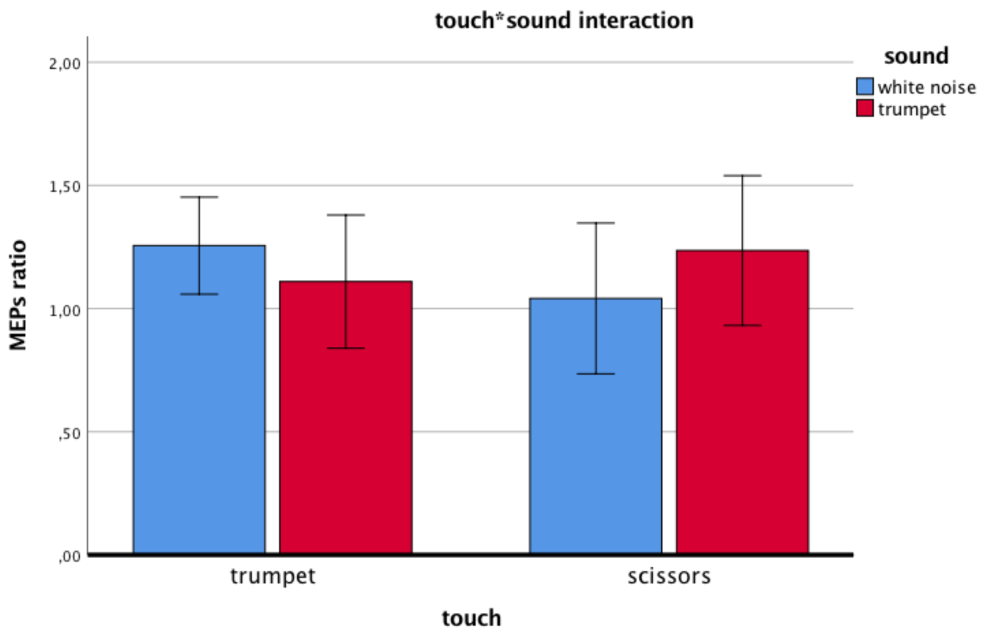
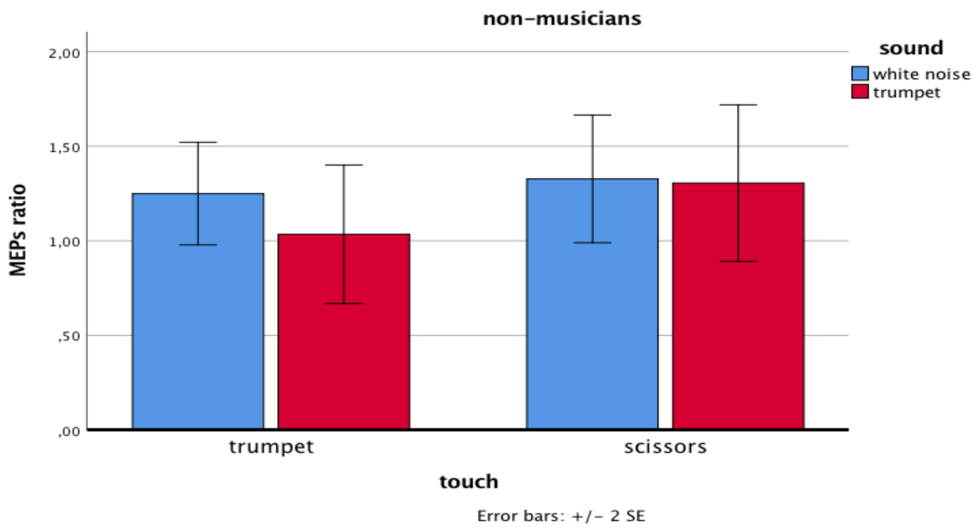


Figure 3. Interaction touch*sound. Despite the significant interaction, no significant differences are revealed by post-hoc comparisons, even if a tendency can be noticed toward higher cortico-spinal excitability when listening to trumpet compared to listening to white noise, while holding a scissors. The opposite pattern is visible while holding the trumpet. Error bars represent the standard error of the mean.



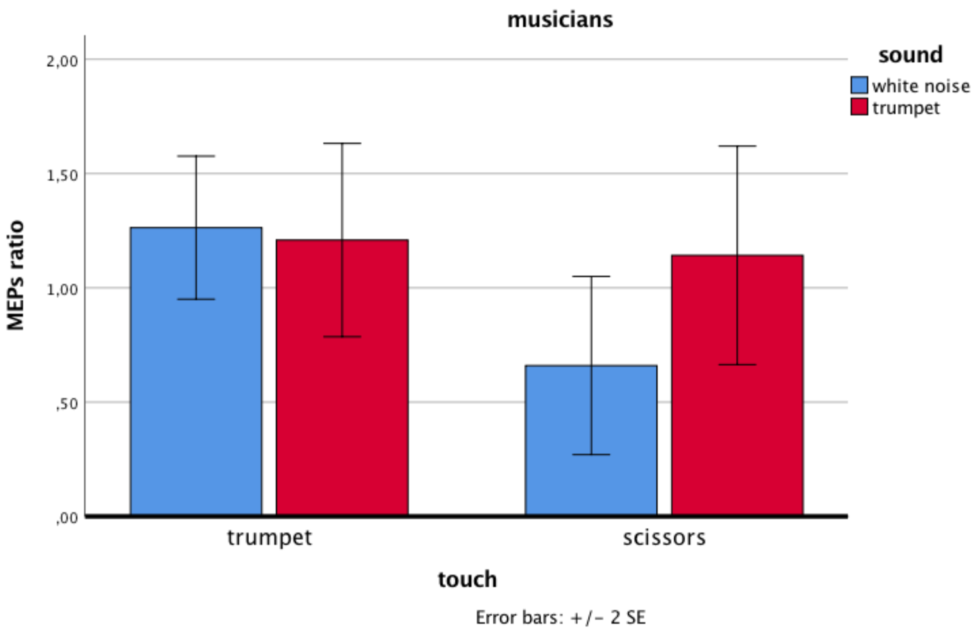


Figure 4. Comparison between experimental conditions. Above. Non-musicians show higher MEPs for white noise while touching the trumpet, but no sound difference while touching the scissors. **Below**. Musicians show the opposite pattern, hence the interaction group*sound and, if we combine both groups, the interaction touch*sound.

Discussion

Playing a wind instrument, like the trumpet, requires the involvement of a complex neuromuscular network including fingers, hands, arms, mouth, tongue and the respiratory system, in tight coordination with the possibilities provided by the musical tool. In the present study we investigated whether and how cortico-spinal excitability of right FDI muscle in wind musicians and non-musicians is modulated by touching a trumpet (compared to a scissors), while listening to a trumpet tone (compared to white noise). We did not find the expected interaction between tool, sound and group, but, in partial accordance with our hypotheses, we found an interaction between group and sound, indicating that musicians tend to exhibit lower MEPs than non-musicians while listening to white noise, independently of the tool held. Moreover, we found a significant interaction between touch and sound, with a trend toward a higher cortico-spinal activation in both musicians and non-musicians, but, rather surprisingly, only in one of the incongruent conditions, that is, when they held the scissors while hearing a trumpet tone.

Although the expected interaction between group, touch and sound was not found, the interaction between group and sound suggests that in this experiment musicians were less sensitive than non-musicians to the white noise, while both groups reacted similarly to the trumpet sound, showing a higher cortico-spinal activation than the baseline in the latter case. The similar cortico-spinal excitability to the trumpet sound in both groups does not seem compatible with the functioning of the so called "echo neurons" (Kohler et al. 2002, see also Gazzola et al. 2006), that part of the mirror neuron system elicited when listening to sounds produced by motor actions that are part of one's repertoire. Indeed, it is well known that the macaque monkey pre-motor cortex is endowed with audio-visuo-motor neurons that discharge not only when the monkey carries out a motor action and when the monkey sees someone doing it (e.g. breaking a nut), but also when it hears the sound produced by the same action (e.g. the sound of a breaking nut). However, as said, we found similar cortico-spinal excitability in both musicians and non-musicians, the latter having no competence in playing music by definition. Nevertheless, the fact that musicians show lower MEPs than non-musicians while listening to white noise may be explained exactly by their higher expertise, which may allow for them to ignore sounds that are not musically relevant. The different neural sensorimotor connectivity between musicians and non-musicians has been quite studied in the last decade. For example, in a MEG study (Lappe et al. 2008), in which non-musicians were assigned either to a group that practiced simple chords at the piano or to a group that simply listened to those very chords, the former group showed higher mismatch negativity than the latter (in the left STG), when presented with a sequence of deviant chords, indicating that the audio-motor practice enhanced the sensitivity to sounds more than simply listening to them (see also Jäncke 2012 and Lee et al. 2011). Moreover, it is likely that musicians exploit a sensorimotor mirror network both when listening to musical excerpts (Herholz & Zatorre 2012) and when observing music-related movements (Candidi et al. 2014). Indeed, the same motor representations required for the actual music production is evoked by those unimodal sensory experiences. This issue can be combined with what we said above about affordances (see *Introduction*), in so far as sounds represent for musicians possibilities to musically interact with other musicians (Windsor & de Bezenac 2012).

The present results extend previous findings by showing that multisensory inputs from somatosensory and auditory stimuli can modulate cortico-spinal excitability of the (index) finger muscles. However, the second of our significant results goes definitely against our prediction. The higher cortico-spinal excitability in one of the incongruent conditions might invoke an explanation related to the surprise induced in a case such as listening to the sound of a trumpet while holding a scissors, given that no sound at all would be expected from a scissors, let alone a trumpet sound. However, if we look deeper into the results of the two groups separately (Figure 4), we can see that non-musicians show higher MEPs for white noise independently of the tool they held, while musicians show the opposite pattern, as was already clear by our first result. Therefore, the related question becomes "why non-musicians were more activated by white noise than by trumpet sounds?" It can be speculated that while the latter do not provide affordances for non-musicians, the former may trigger a sort of startle effect in them, that is, "a rapid involuntary triggering of a prepared movement in response to a loud startling acoustic stimulus" (Smith et al. 2019, Alibiglou & MacKinnon 2013), which, in wind musicians, is overcome by a relevant sound, that is, a sound coming from a wind instrument.

A clear limitation of the present study is the investigation of a muscle, FDI, which is as often used in playing an instrument as it is in many other daily activities, from pressing a button to switch the light on to writing on a computer keyboard or a smartphone. Indeed, we plan to record from a muscle, which is more specifically involved in playing a wind instrument, like orbicularis oris (Gotouda et al. 2007, Potter et al. 2015), though its higher motor threshold requires a careful monitoring of its contraction (Adank et al. 2018). Interestingly, single-pulse TMS on this muscle, suggested a dissociation between singing and speech perception, with the former inducing a cortico-spinal decrease in the left hemisphere, compared to a baseline (and compared to the right hemisphere, Royal et al. 2015). Moreover, perceiving larger sung intervals were associated with higher cortico-spinal excitability than perceiving smaller intervals. Although it remains to be established if such an effect is related to the pitch amplitude of the intervals or to the respective visual input (the subjects of these experiment saw a person either singing intervals or telling short proverbs), one might reasonably conclude that the motor cortex corresponding to the orbicularis oris is a fundamental site to investigate the neural underpinnings of wind instruments playing (see also Gebel et al. 2013 for an fMRI study on trumpet and piano players). Another possible confounding variable in our experiment could be the presence of wind musicians other than trumpet players, like saxophone and trombone players, but the fact that two out of three outliers in the musicians sample were precisely trumpet players should attenuate such a worry.

Conclusions

A variety of studies in the last twenty years have shown that tool use can shape the way we perceive the body space and the space around us. We do not know if the static nature of our paradigm prevented the emergence of such a modulation on cortico-spinal excitability while integrating a tool and its corresponding sound. However, an interaction was found between the kind of tool and the sound subjects listened to, although pointing to the opposite direction of our predictions. More interestingly, the interaction between group and sound confirmed that expertise with a given tool might somehow influence the way the neuro-muscular system code for a sound, even if FDI in wind instruments playing is likely to be the weakest candidate within such a system, compared, for example, to orbicularis oris or other facial muscles.

References

Adank P, Kennedy-Higgins D, Maegherman G, Hannah R & Nuttall HE (2018). Effects of Coil Orientation on Motor Evoked Potentials From Orbicularis Oris. *Frontiers in Neuroscience* 12:683. doi: 10.3389/fnins.2018.00683

Alibiglou & MacKinnon (2012) The early release of planned movement by acoustic startle can be delayed by transcranial magnetic stimulation over the motor cortex. *Journal of Physiology*, 590.4 pp 919–935

Baily, J. (1992). Music performance, motor structure, and cognitive models. In M. P. Baumann (Ed.), *European studies in ethnomusicology: Historical developments and recent trends* (pp. 142–158). Wilhelmshaven, Germany: Floian Noetzel Verlag.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H. & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, 30, 917–926.

Berti, A., & Frassinetti, F. (2000). When far becomes near: Remapping of space by tool use. *Journal of Cognitive Neuroscience*, 12, 415–420. https://doi.org/

10.1162/089892900562237

Brown LE. & Goodale MA. (2013). A brief review of the role of training in near-tool effects. *Frontiers in Psycholo*gy, **4**:576. doi: 10.3389/fpsyg.2013.00576

Buick AR., Kennedy NC. & Carson RG. (2016). Characteristics of corticospinal projections to the intrinsic hand muscles in skilled harpists. *Neuroscience Letters* 612, pp. 87–91

Candidi M, Sacheli LM, Mega I & Aglioti SM. (2014) Somatotopic mapping of piano fingering errors in sensorimotor experts: TMS studies in pianists and visually trained musically naıves. *Cerebral Cortex* 24, 435 – 443. (doi:10.1093/cercor/bhs325)

Clark A. (2003). *Natural Born Cyborgs*, Oxford University Press, Oxford

Clark A. (2008). *Supersizing the Mind*, Oxford University Press, Oxford

Clark A. & Chalmers D. (1998). The Extended Mind, *Analysis*, Vol. 58, No. 1, pp. 7-19

Dell'Anna A., Fossataro C., Burin D., Bruno V., Salatino A., Garbarini F., Pia L., Ricci R., Leman M. & Berti A. (2018) Entrainment beyond embodiment, *Neuropsychologia* 119, 233–240

Dell'Anna A., Rosso M., Bruno V., Garbarini F., Leman M & Berti A. (forthcoming), Does a musical interaction in a jazz duet modulates peripersonal space? *Psychological Research*

Elbert T., C. Pantev, C. Wienbruch, B. Rockstroh & E. Taub, (1995). Increased cortical representation of the fingers of the left hand in string players, *Science* 270 (5234) 305–307.

Fossataro, C., Bucchioni, G., D'Agata, F., Bruno, V., Morese, R., Krystkowiak, P. & Garbarini, F., (2017). Anxiety-dependent modulation of motor responses to pain expectancy. *Social Cognitive Affective Neuroscience*. Vol. 13, N. 3, pp. 321–330.

Gazzola V, Aziz-Zadeh L. & Keysers C. (2006) Empathy and the somatotopic auditory mirror system in humans. *Current Biology* 16, 1824 – 1829. (doi:10.1016/j. cub.2006.07.072)

Gebel B., Ch. Braun, E. Kaza, E. Altenmüller & M. Lotze (2013) Instrument specific brain activation in sensorimotor and auditory representation in musicians, *Neuroimage,* 74:37-44. doi: 10.1016/j.neuroimage.2013.02.021.

Gentner R., S. Gorges, D. Weise, Kaufm Kampe, M. Buttmann & J. Classen (2010) Encoding of motor skill in the cortico-muscular system of musicians, *Current Biology* 20 (20) 1869–1874.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gotouda A., T. Yamaguchi, K. Okada, T. Matsuki, S. Gotouda & N. Inoue (2007) Influence of playing wind instruments on activity of masticatory muscles. *Journal of Oral Rehabilitation* 34; pp. 645–651

Haueisen J & Knosche TR. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience* 13, pp. 786–792. (doi:10.1162/089892901525441449)

Herholz, S. C. & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behaviour, function and structure. *Neuron,* 76, pp. 486–502.

Hunley SB. & Lourenco SF. (2018). What is peripersonal space? An examination of unresolved empirical issues and emerging findings. *Wiley Interdisciplinary Review Cognitive Science;* 9: pp. 1–17. doi: 10.1002/wcs.1472

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body-schema during tool use by macaque post-central neurones. *Neuroreport*, 7, 2325–2330.

Jäncke, L. (2012). The dynamic audio-motor system in pianists. *Annals of the New York Academy of Sciences*, 1252, 246–252.

Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., & Pantev, C. (2014). Audio-tactile integration and the influence of musical training. *PLoS One*, 9, e85743.

Landry S.P. & Champoux F. (2017). Musicians react faster and are better multisensory integrators. *Brain and Cognition* 111, 156–162

Lappe C., Herholz SC., Trainor L.J. & Pantev C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *Journal of Neuroscience*, 28, pp. 9632-9

Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences*, 108, E1441–E1450.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MIT Press

Maravita, A. & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, 8, 79– 86. https://doi.org/10.1016/j.tics.2003.12.008

Menary R. (2010) ed. *The Extended Mind*, MIT Press, Cambridge (MA)

Munte, T., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3, 473–478.

Osborne, J. (2002). Notes on the use of data transformations. Practical Assessment, Research & Evaluation, 8(6), 1–7. Retrieved from http://pareonline.net/getvn.asp?v=8&n=6

Patané, I., Iachini, T., Farnè, A., & Frassinetti, F. (2016). Disentangling action from social space: Tool-use differently shapes the space around us. *PLoS One*, 11(5), e0154247.

Potter N., Lauren R. Johnson, Stephen E. Johnson & Mark Van Dam (2015). Facial and Lingual Strength and Endurance in Skilled Trumpet Players. *Medical problems of performing artists*, 30 (2), pp. 90-95

Ranganathan R., Lee MH., Brown A. & Newell K. (2011). Grasping possibilities for action: influence of object function and action capabilities. *Human movement science* 30(6):1102-14

Reybrouck, M., Vuust, P., & Brattico, E. (2019). Music and the plastic brain: How sounds trigger neurogenerative adaptations. *Neuroplasticity*. Rijeka, Croatia: InTech

Royal I., Pascale L., Théoret H., Russo F.A. & Peretz I. (2015) Excitability of the motor system: A transcranial magnetic stimulation study on singing and speaking, *Neuropsychologia* 75, pp. 525–532

Sartori, L., Straulino, E. & Castiello, U. (2011). How objects are grasped: the interplay between affordances and end-goals. *PLoS One*, 6, e25203.

Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. *Progress in Brain Research, 217*, 37-55. doi:10.1016/bs.pbr.2014.11.020

Schulz, M., Ross, B. & Pantev, C. (2003). Evidence for training-induced cross modal reorganization of cortical functions in trumpet players. *NeuroReport*, 14, 157–161.

Serino A., Bassolino, M., Farnè, A., & Làdavas, E. (2007). Extended multisensory space in blind cane users. *Psychological Science*, *18*(7), 642–648.

Shaw, R., & Turvey, M. T. (1981). Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy, & J. Pomerantz (Eds.), *Perceptual organization* (pp. 343–415). Hillsdale, NJ: Lawrence Erlbaum.

Smith V, Maslovat D., Carlsen AN. (2019) StartReact effects are dependent on engagement of startle reflex circuits: support for a sub-cortically mediated initiation pathway. *Journal of Neurophysiology*, 122(6): 2541-2547

Windsor W. & de Bézenac C. (2012). Music and affordances, *Musicae Scientiae*, *16*, 102-120

Yamaguchi K., Nakamura K., Oga T. & Nakajima Y. (2014) Eating tools in hand activate the brain systems for eating action: A TMS study, *Neuropsychologia* 59, 142–147

Chapter 5. Timing markers of interaction quality during semi-hocket singing

Dell'Anna A.*[5] [6] , Buhmann J. [5], Six J. [5], Maes PJ. [5], Leman M. [5]

Abstract

Music is believed to work as a bio-social tool enabling groups of people to establish joint action and group bonding experiences. However, little is known about the quality of the group members' interaction needed to bring about these effects. To investigate the role of interaction quality, and its effect on joint action and bonding experience, we asked dyads (two singers) to perform music in medieval "hocket" style, in order to engage their co-regulatory activity. The music contained three relative inter-onset-interval (IOI) classes: quarter note, whole note, and dotted whole note, marking time intervals between successive onsets (generated by both singers). We hypothesized that singers co-regulated their activity by minimizing prediction errors in view of stable IOI-classes. Prediction errors were measured using a dynamic Bayesian inference approach that allows us to identify three different types of error called fluctuation (micro-timing errors measured in milliseconds), narration (omission errors or misattribution of an IOI to a wrong IOI class), and collapse errors (macro-timing errors that cause the breakdown of a performance). These three types of errors were correlated with the singers' estimated quality of the performance and the experienced sense of joint agency. We let the singers perform either while moving or standing still, under the hypothesis that the moving condition would have reduced timing errors and increased We-agency as opposed to Shared-agency (the former portraying a condition in which the performers blend into one another, the latter portraying a joint, but distinct, control of the performance). The results show that estimated quality correlates with fluctuation and narration errors, while agency correlates (to a lesser degree) with narration errors. Somewhat unexpectedly, there was a minor effect of movement, and it was beneficial only for good performers. Joint agency resulted in

[5] IPEM, Ghent University, Belgium
[6] Department of Psychology, Turin University, Italy

a "shared", rather than a "we", sense of joint agency. The methodology and findings open up promising avenues for future research on social embodied music interaction.

Introduction

Music is a rewarding and empowering activity (Chanda & Levitin 2013, Fritz et al. 2013), having the capacity to connect people (Malloch & Trevarthen 2009, Overy & Molnar-Szakacs 2009) and increase their self-confidence, their feelings of wellbeing, for example after singing together (Kreutz 2014), and their motivation, for example in treating neurological disorders (Wan et al. 2010, MacDonald et al. 2012). While there exist other such facilitators of reward and empowerment, such as dance, ritual actions, sports, and other forms of joint actions (Sebanz et al. 2006), music is special in the sense that its social power is driven by auditory information, next to visual information. When you don't see what the other is doing, your action can still be perfectly synchronized. And when muscles and brains get synchronized, strong group bonding effects may occur (McNeill 1995). Obviously, while the use of music may date back to the very beginning of human evolution (Honing et al. 2015), its power is still working in all kinds of human social activities, including academic meetings, banquets, funerals, rituals, football matches, concerts, festivals, and so on.

Music can drive joint actions, intended as a "form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz et al. 2006: 70), and generate affects (e.g. of being in joint control and connected with others, see Keller et al. 2016 for a review). According to the minimal architecture hypothesis (Vesper et al. 2010), joint action can be investigated in terms of representations (the goal and tasks the subjects involved in it assign themselves), processes (prediction and monitoring of the various steps needed to accomplish it) and coordination smoothers (actions that simplify coordination). In this paper, we focus on timing prediction as a dynamic marker of the quality of a musical joint action in singing dyads and correlate it to subjective reports of that very quality and of the joint agency induced by it.

In recent works, advantage has been taken of imposed musical tasks in order to understand how two or more subjects build representations of a joint action, employing both behavioural (Keller et al. 2007, Goebl & Palmer 2009, Lesaffre et al. 2017) and neuroscientific (Loehr et al. 2013, Keller et al. 2014, Arbib 2013) approaches to better understand music-based social empowerment (D'Ausilio et al. 2015). In a couple of works (Müller & Lindenberger 2011, Müller et al. 2018), it has been shown that singing together (in a choir) implies also breathing and heart beating together, giving rise to a complex network of processes that, with the addition of body movements, imposes boundary conditions to its constituents (the singers), just like a "superordinate system". In other words, singing together consists of a "participatory sense-making" that spreads out in the dynamics of the interaction itself, back to the subjects who get individually affected by certain properties of the singing (De Jaegher & Di Paolo 2007, Schiavio & De Jaegher 2017). Thereby, interaction cannot be understood by analysing one single subject at a time, but rather by analysing the interaction itself (in the form of behaviour relative to one another).

However, at this point the question arises as to how good a music interaction should be in order for the music's "bio-technological power" (Freeman 2000) to become effective. A joint action such as singing can facilitate group-formation and generate feelings of connectedness, yet to what degree is this feeling depending on interaction qualities, that is, on the capacity to perform the rules as stated by the cultural context? In our opinion, few studies have addressed this question. Recent studies, indeed, use techniques that measure the bodily interaction of musicians mostly in view of timing, but quality as such is not addressed (Loehr et al. 2009, Eerola et al. 2018). To this effect, quality can be estimated through self-assessments by performers or third persons. However, it is also of interest to consider we exploited the concept of joint agency. Pacherie (2012), in generalizing the concept of agency (Haggard & Eitam 2015) from the individual to the group, distinguishes between a SHARED sense of joint agency and a WE sense of joint agency, pointing out that, if an action is a joint action, the resulting sense of agency should be a feeling of being in control of (at least) part of the joint action outcome. According to Pacherie's distinction, people may experience a SHARED sense of joint agency in small groups, with a certain degree of specialization among the different participants, but without hierarchies, while a WE-agency might be experienced in larger ensembles with less specialization among its members and (sometimes) directed by a leader. Fitting examples are a volleyball team for the former kind, and a volleyball stadium choreography for the latter and, from a musical point of view, a small combo and an orchestra, respectively. In both cases, Pacherie stresses the importance of predictability, but to a different degree: a SHARED sense of joint agency may draw upon a low predictability of the partners' action, whereas a WE-agency may draw upon a high predictability due to similarity among partners' actions (what she calls "coordination symmetry"). Pacherie's model has been successfully applied to dyads in studies comparing individual and shared control on a given joint action (Bolt et al. 2016, Bolt & Loehr 2017). We extend this investigation so to include the WE-agency factor, trying to establish whether the (two, in our experiment) performers experience distinction from each other or blending into each other. The latter would imply a kind of boundary loss between agents.

To address the analysis of timing, we adopt a Bayesian inference framework to guide our methodological choices. Bayesian inference is the core approach behind the predictive coding theory (Friston 2010, Clark 2016, Vuust et al. 2009, Koelsh et al. 2019), a nowadays largely debated theory that sees the brain as an active generator of predictions, rather than a passive receptor of stimuli from the external world. Thanks to sensory feedback received in a continuous circular sensorimotor process, prediction errors are minimised for a given action the subject is about to accomplish. Brain networks thus formulate hypotheses about the possible state of the world (the prior) that are compared to the actual sensory information received (the error), in order to update the hypothesis (the posterior). Priors and posteriors are also called "beliefs", not in the sense of explicit propositions, but rather in the sense of probability distributions, hence, mainly latent variables. A continuous updating of such beliefs is allowed by acting on the world in ways that minimize the error, or sensory surprise. Such sensorimotor loops, then, work as "active inferences" (Adams et al. 2015).

In a social context of music interaction, feedback on timing is provided by the other interacting subjects' behaviour. As stressed by Koelsch et al. (2019), music is a perfect case against which the predictive model may be tested, because the music's very syntactic structure implies rhythmic, melodic and harmonic expectancies, that is, prediction. We believe that accurate prediction of the partner's action is crucial in musical ensembles, even if it shows in different degrees, depending on musical genres, cultures and kind of ensembles (Rohrmeier & Koelsch, S. 2012, Salimpoor et al. 2015). In this paper we focus on timing since this is one of the main features in which the quality of music is reflected, and probably the most tractable with our approach. The Bayesian inference framework is here used to develop a computational analysis that copes with prediction errors in conditions where timing can be unstable, thereby assuming that performers construct latent time-varying beliefs about their joint timing. Our quest for interaction quality is therefore also a quest for a proper methodology for estimating latent variables about joint timing. Importantly, such a timing marker has to be considered as a dynamic index of coordination insofar as it takes into account not only the timings of the two musicians separately and correlate them afterwards (for example, by means of windowed cross-correlation), but several inter-onset intervals constituted by the two interacting musicians' singing (see Methods). It is worth stressing that, while timing errors may be due to a variety of factors (inability of reading the notes, general lack of ability to accurately follow to synchronize singing with beat, etc.), in this paper we are interested in the dynamics, rather than the causes, of timing quality.

Following the growing interest in embodied approaches to cognition (Thompson & Varela 2001, Gallagher 2005, Chemero 2009), in particular music cognition (Iyer 2002, Leman 2007, Walton et al. 2015), the kinaesthetic dimension of musical performances has been widely explored in the last years, stressing the impact of body movements on both production and perception. The predictive coding framework, combined with these embodied approaches (Gallagher & Allen 2016), can help explain how movement, along with other sensory modalities, could contribute to error minimization. Indeed, when body parts move in time with the music, their timing reflects the timing of the music and can help shape this timing during production (be it singing or playing an instrument, see Wanderley et al. 2005, Maes et al. 2014).

In the present study, we explore interaction quality in a singing dyad, taking advantage of the medieval "hocket" style, in which two (or more) musicians are required to build a melody together using strict alternation of notes. Our hypotheses are as follows:

(i) The quality of music interaction is reflected in the timing of the joint action among performers. Performers can estimate their own interaction quality through continuous video annotation (video-stimulated recall) and they can assess the social effect of the interaction in terms of an assessment of their joint agency experienced during the interaction. As quality is reflected in timing, joint action timing can be measured as the performers' latent (or emerging) belief about joint timing. We predict that more accurate timing is correlated with higher quality, as reflected (i) in the performers' higher self-annotation of their own performed interaction quality, and (ii) in the performers' estimation of joint agency experiences.

(ii) Given the high similarity of the singers' music score (see Figure 1), a high quality in performing will correspond with a high sense of joint agency values, that is, by WE-agency.

(iii) Movement may help performers to make their timing more accurate. Indeed, since multiple senses take away uncertainty (according to the predictive coding theory) and movement is timing (according to embodiment theory), movement should affect quality.

Materials and methods

Ethics statement

Participants were informed in advance about the task, the procedure and the technology used for measurement. They had the opportunity to ask questions and were informed that they could stop the experiment at any time. The ethics committee of the Faculty of Arts and Philosophy of Ghent University approved the study and the consent procedure.

Participants

Fifteen couples of musicians were recruited (mean age 29.4 ± 10.4 years; 12 women), both participants being either men or women, so that their pitch range could match more easily. As musicians we considered people currently playing an instrument (or singing) with at least five years of regular (formal or informal) musical training (mean 10.1 ± 9.7), capable of singing a simple melody from sheet music.

Task

In this experiment we let pairs of musicians sing "on stage" an interleaved melody provided on a score. They were told that their parts should never overlap, and that the combination of the parts would result in a melody consisting of an A- and a B-part. They were also instructed to try to keep going if for some reason their interactive performance would break up. We asked the participants to sing the notes by producing the sound "ta" or "pa". The fact that these sounds start with a plosive facilitates automatic onset detection of notes, needed to extract inter-onset-interval (IOI) durations. In hocket polyphonic style a single melody is broken into two or more parts that never overlap, alternating almost regularly one tone after another. Here we use a semi-hocket technique for two singers, where alternation is somewhat less strict, meaning that sometimes a singer might sing two notes in a sequence (see score, Figure 1). Obviously, the quality of the singing can be assumed to be reflected in the performers' timing, and thus on the contribution of each partner to the whole. Due to a limited rehearsal time (5 min alone, 15 min together) the task was expected to be challenging, leading to different outcomes in performance quality. After the rehearsal, singers had to perform eight trials of two randomized conditions lasting two minutes each, either moving (four trials) or not moving (four trials). In the non-movement trials participants were asked to stand as still as possible, while performing the singing task. In the movement trials participants were invited to move as they pleased while performing. This could result in simple hand- or foot-tapping, head-nodding, body-swaying, or even dance-like movements.

Technical setup

For each recording of the musical interaction task the two participants were standing on a force plate facing each other. Both force plates have four weight sensors at the corners in order to register movement of the participants. The measured voltages are converted to MIDI CC messages by means of a Teensy 3.2 microcontroller. The MIDI stream was recorded in Ableton 9. The encoding of sensor signals into MIDI makes it straightforward to record audio and sensor data in sync using standard DAW software such as Ableton. A decoder script turns the MIDI into data fit for analysis. Participants were equipped with a headset containing a small microphone. The singing was thus captured and also recorded with Ableton. In addition, a video recording was made with a webcam (Logitech, c920). The webcam was modified to allow audio input. The audio input is connected to a SMPTE source to synchronize the video with the audio.

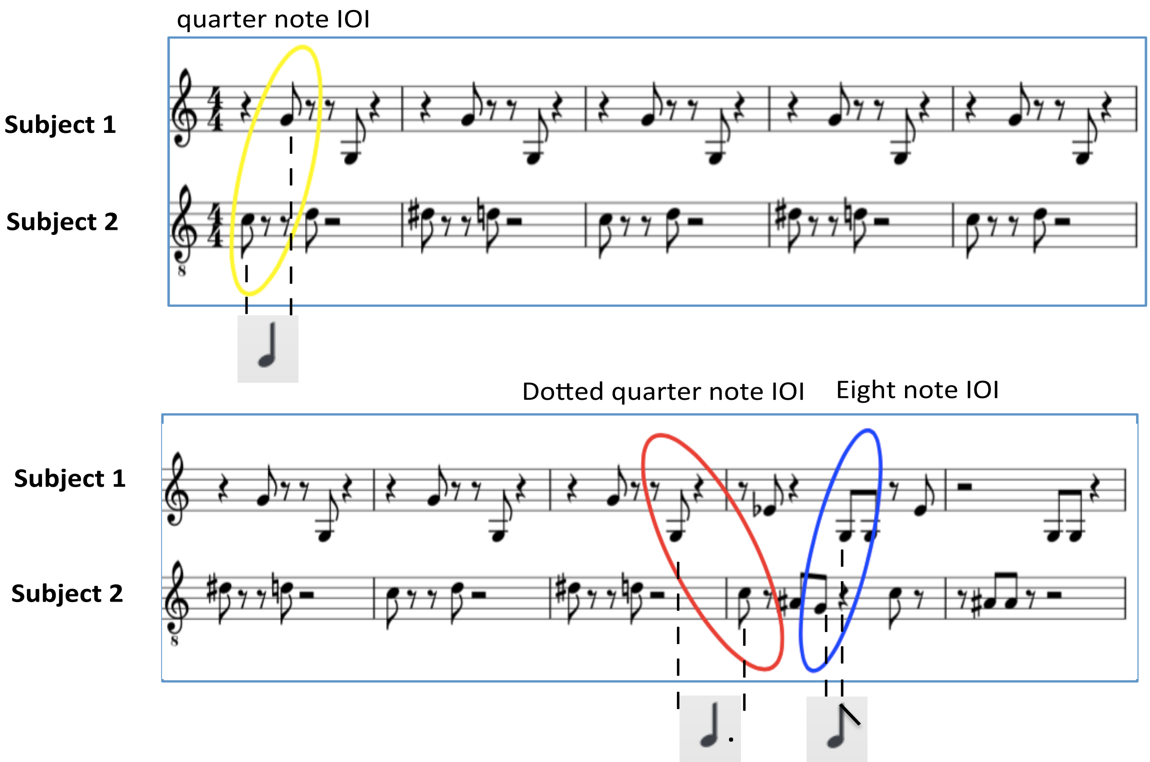


Figure 1. **Experimental stimulus**. Part of the participants' scores. Together these scores form a semi-hocket, meaning that there are no simultaneous notes and the combined scores merge into one melody. This melody is an adaptation of Michael Jackson's Billy Jean. The two parts contain an equal number of notes, displaying the same level of difficulty. In yellow, red and blue the three IOIs used in Bayesian regression (see below) are highlighted

These audio-visual recordings were used immediately after the recording session. The participants were requested to review their performances and annotate the quality level of their interaction. This was done via a script that synchronized and merged the audio and video recordings per trial. The scoring of the interaction happened on two separate computers via the mouse that could move a visual line up (better quality) and down (worse quality). The visual scores are stored as thousand samples of values between 0 and 127. The initial position of the cursor was set to value 64, a neutral starting point. All recording devices were connected with a master sync clock (Rosendahl Nanosyncs HD), preventing drift and enabling precise synchronization of audio movement, and video data.

Procedure

Each couple was welcomed in our laboratory and, after filling in the informed consent, participants were explained that they had to build a melody together, combining their individual parts, stressing that these should never overlap, but almost always alternate one note after the other. Moreover, subjects were not allowed to read their partner's score. Then, they rehearsed their part in two separate rooms for 5 minutes, having the opportunity to listen to it once or twice in order to find the right pitch and learn the melody. Afterwards, they were gathered in the main lab, equipped with the headsets, and invited to get on "the stage", that is, on the two balance boards facing each other at 1.5 m, to rehearse together for 15 minutes maximum, before the beginning of the real performance. After these recordings, each participant individually executed a quality assessment task concerning the performance and the sense of joint agency (see below), without communicating with the partner. In total the experiment took between 1.5 and 2 hours per couple.

Data pre-processing

Audio onset-detection: As our approach has a focus on timing, the audio recordings are reduced to the onsets of the singing, by doing an automatic onset detection in *Sonic Visualizer* followed by a manual checking and correction step. These onsets are converted into IOI durations. Our analysis is based on the relative IOIs, that is, the IOIs formed by the alternated singing of the performers. This implies that, overall, a relative IOI is defined by the onset of one performer and the onset of another performer, except in the cases where a performer sings two consecutive notes. In that case the IOI is defined by the consecutive onsets of the same performer. In theory or in case of an accurate performance, this results in three types of IOI durations, matching the durations of eighth notes, quarter notes, and dotted quarter notes (Figure 1). Depending on the tempo, a 2-minute performance equals approximately singing the A- and B-parts four times. In theory this would result in 176 eighth notes, 96 quarter notes, and 47 dotted quarter notes.

Measurable markers of interaction quality

Annotation and questionnaires: for each couple the two participants were asked to assess the general quality of the interaction, that is, the performance as a whole, rather than the quality of their performance, or the other participant's performance. This resulted in two time-series of quality values between 0 and 127 for each trial. Secondly, participants were asked to assess the joint sense of agency on a 7-point Likert scale. In particular, for each of the eight trials the subjects were asked to answer the question "When looking at the moments with the highest quality assessment, how was your feeling of control over the process on a scale between 0 (independent), 3 (shared) and 6 (complete unity with your partner)?" We explained this question by saying that the interaction could be either the product of two actions not really well coordinated between them (independent) or the product of two coordinated but distinct actions (shared) or the product of two actions that are not felt as different, but rather as the accomplishment of a single subject.

Third-person quality assessment: Given the fact that there was a large variation in performance quality, the authors of the paper agreed upon a subjective classification of the performances per duo into two groups, i.e. expert group and non-expert group. This was done by looking at the performance videos and evaluating the stability of the performance (could couples keep up their performances without too many break-ups) and how similar the performance was to what was written in the score. Six couples were assigned to the non-expert group and nine couples to the expert group. This subjective classification was done to validate our hypothesis that a good performance has less performance errors than a bad performance (Figure 3).

Performance errors: The score defines a musical norm for interactive performances, including rhythmic figures, tempo and an overall melodic narrative. However, due to the fact that the music emerges from the interaction, we assume that singers predict each other's performance in order to perform their own contribution correctly. As mentioned, not all performances may reach a high-quality level of interaction. Given the constraints of the musical rules, we consider three different types of prediction-errors, related to:

- *Fluctuation*: The fluctuation errors are defined as micro-timing (in milliseconds) prediction-errors that result from different sources such as timing-corrections due to small mistakes, due to active sampling, or even small onset measurement errors within the data pre-processing. Overall, fluctuation is a source of variance that can be considered necessary in order to maintain a stable performance state, even of high quality.
- *Narration*: The narration errors are defined as meso-timing (typically up to half a second, related to note durations) prediction-errors that may occur when a performer fails to follow the musical rule, for example, by forgetting a note, or making a mistake in note duration. Pitch is not taken into account. Overall, an error in the sequence (for example due to the omission of a note) may disturb the ongoing interaction. However, the dynamic system may be resilient enough to recover from such errors.
- *Collapse*: The collapse errors are defined as macro-timing (up to several seconds) prediction-errors that may occur when the performance, hence also the musical interaction, breaks down. The breakdown is catastrophic in the sense that both performers lose control of the expected musical narrative. This error is different from the narration errors that allow recovery due to resilience. To recover from such an interaction collapse, it may be necessary to start a new narrative from the beginning of the piece or the beginning of a section.

Data analysis

Bayesian inference approach: As our data-analysis approach is based on the idea that performers try to reduce performance errors with respect to predictions, we consider performers as components of an interaction dynamics. We assume that each performer makes a prediction of the timing of the joint action (the interaction) based on a latent, or emergent, variable that estimates the timing of the relative IOIs in terms of milliseconds. As the piece contains only three different IOI classes, we assume that performers construct a latent variable for the estimated timing of each IOI-class. Obviously, the timings of the IOI classes are mutually constrained, thus contributing to a global latent variable, which is known as tempo. In our analysis we focus on how performed IOIs relate to the latent IOI-classes. Rather than inferring the prediction errors from an estimated global tempo (and proportional ratio of that tempo with respect to the IOI-classes) our method is tolerant to a systematic shortening or lengthening of IOI-classes according to performers' expressive timing preferences. The initial values of the variables that estimate the timing of the IOI-classes are set by a k-means clustering on all IOIs in three IOI-classes, using the first 15 seconds of a performance. Thereafter, a sequential Bayesian updating is performed for each of the IOI-classes separately, using a 15-seconds window of incoming IOI values (leading to the evidence distribution or the likelihood of measurement). Using Bayesian terminology, we interpret the prior as the mean of a distribution of old predicted durations of the IOI-class and the posterior as the mean of an updated distribution due to new evidence. This procedure is executed step by step (i.e. one IOI after another in the time series). It allows us to calculate the difference between the performed IOI and the predicted IOI, in milliseconds (Figure 2). For the entire performance, we calculate the root-mean-square error (RMSE) for each IOI-class, and take the average over all IOI-classes. This approach can deal with small changes in tempo and therefore, it accounts for the assumption of non-stationarity. In fact, for each IOI-class we use proportional timing errors by taking the log2 of the ratio of the measured IOI and the predicted IOI.

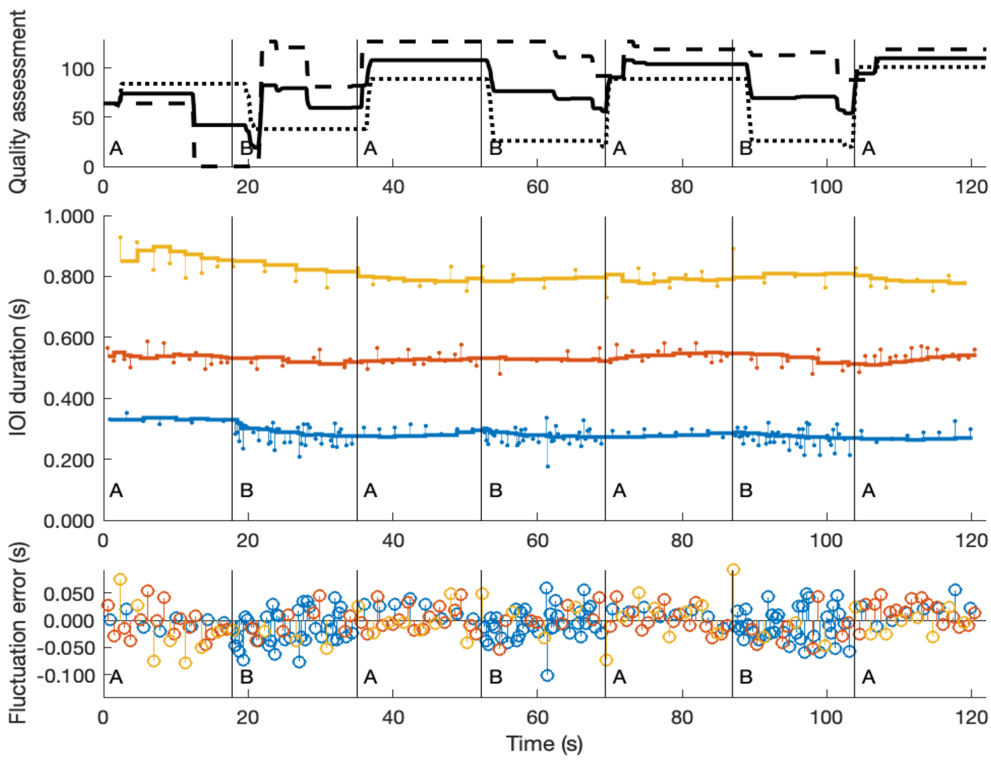


Figure 2. Performance interaction measurements in time-series. In this figure, three time-series of the same performance (a 2-minute trial) are visualized. The top plot shows the two quality assessments of the two participants (the dotted lines); the solid line represents the mean of the two assessments. The horizontal lines of the middle plot show how the priors of the three IOI classes evolve over time, according to our Bayesian sequential updating approach. The vertical lines with small dots indicate the deviations of the actual IOI durations with respect to those priors; in other words, they represent the errors in seconds. The plot at the bottom is a summary of the middle plot, where the zero-line represents the priors and the vertical lines are the errors (in seconds) for the three IOI classes (blue = short IOIs; orange = middle IOIs; yellow = long IOIs).

While the above approach may be working for fluctuation errors, we also have to consider the fact that IOIs may be wrongly classified due to narration errors. In order to account for these narration errors (which are restricted to duration errors), we keep track of the sung (duration) sequence and the expected corresponding IOI-class assignments. When expected IOIs get wrongly classified they are considered as narration errors, expressed in percentage of matching IOIs. Collapse errors are considered to be larger gaps in the performance (IOI durations that differ more than two standard deviations from the corresponding IOI-class prior), where normally onsets would have been expected. The collapse errors are expressed as a percentage of the number of collapses compared to the total number of IOIs in the performance.

Correlations between first-person viewpoints and timing errors: The correlation was calculated between the overall timing errors per trial and the average joint-agency scores that were indicated in the questionnaires. This was done for each performance-error type. Since the joint agency scores are not normally distributed and they contain a lot of identical values (7-point Likert scale), Kendall's Tau correlation was used.

The effect of movement and expert group on performance errors: In order to test our hypothesis that movement has an auxiliary function in error minimization during a joint singing task, for each type of performance error (fluctuation, narration and collapse) we compare the average error value of the four movement trials with the average of the four non-movement trials. A 2 x 2 mixed ANOVA was performed with condition (movement/non-movement) as within-subject factor and expert-group (yes/no) as between-subjects factor. In a few cases the performance errors in a group were not normally distributed. Non-parametric tests were executed as well (Mann-Whitney U tests to compare experts with non-experts for each condition; Wilcoxon signed rank tests to compare movement with non-movement condition for each expert level), revealing similar results to the ANOVA. To maintain uniformity, we report solely the parametric tests.

The effect of movement and expert group on agency and quality assessment: To validate our hypothesis that movement and expert level have a positive impact (higher agency and quality scores for experts, while moving) on the subjective assessment of a performance interaction, a 2 x 2 mixed ANOVA was performed. Identical to the test on performance errors, condition (movement/non-movement) is the within-subject factor and expert-group (yes/no) the between-subjects factor.

Movement assessment: For each trial, continuous wavelet transforms were performed on the movement data of the two force plates, i.e. for each force plate the sensor that captured the highest amplitude. Only the wavelet information within the movement-relevant range of 0.25 to 5 Hz was considered. Within that range the frequency band with the highest average wavelet magnitude was selected. For each force plate this average magnitude was used to calculate the average movement magnitude for the couple. The right-skewed histogram of these values for all the non-movement trials covers a small range of magnitude values, with a maximum average magnitude value of 11. For the movement trials the histogram covers a much wider range of magnitude values, with a maximum of 88. In accordance with what was observed in the video recordings, a threshold of 25 was chosen as the cut-off for detected movement (above) or not (below).

Results

Effect of movement and expert group

Performance errors: In total, nine out of the 120 performance trials (7.5%) were excluded from analysis. Two trials were excluded (the first and third trial of duo 5), because the participants made a lot of errors by singing (almost) simultaneously, resulting in IOIs that were too scattered and different from what we expected from the performance. Seven more trails were excluded (duo 5, trial 2; duo 12, trial 1; duo 13, trial 2 and 4; duo 18, trial 3; duo 19, trial 5; duo 21, trial 7), because too much movement was detected in the conditions where participants were instructed not to move.

Fluctuation errors are not significantly lower in the movement condition (M = 0.185, SE = 0.022) than in non-movement condition (M = 0.217, SE = 0.036), $F(1, 13)$ = 3.929, p = .069, r = .48. There was a significant effect of expert level, indicating that experts had lower error rates (M = 0.133, SE = 0.012 for movement; M = 0.147, SE = 0.013 for non-movement) than non-experts (M = 0.245, SE = 0.036 for movement; M = 0.298, SE = 0.064 for non-movement), $F(1, 13)$ = 7.938, p = .015 , r = .62. No significant interaction effect was found between movement and expert level, $F(1, 13)$ = 1.436, p = .252, r = .32.

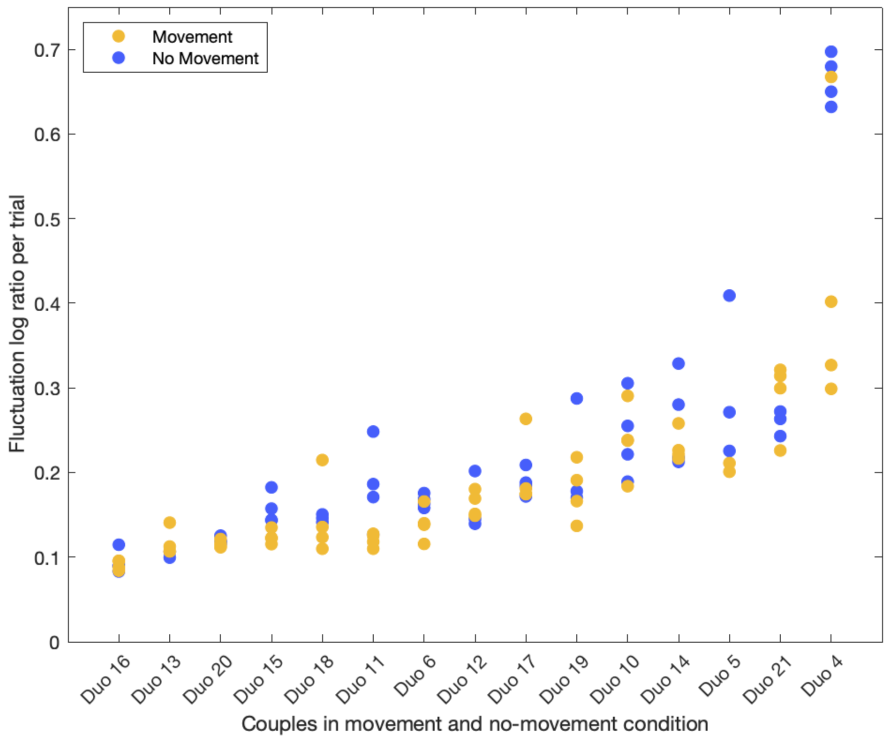


Figure 3. Fluctuation errors per duo. The duos are ordered from smallest to largest average fluctuation error over the eight performance trials.

Narration matching in the movement condition (M = 84.54, SE = 3.43) is not significantly different from that in the non-movement condition (M = 82.90, SE = 3.69), $F(1, 13) = 1.437$, $p = .252$, r =.32. There was a significant effect of expert level, indicating that experts had higher percentages of predictable IOI classes (M = 94.92, SE = 1.44 for movement; M = 93.18, SE = 1.98 for non-movement) than non-experts (M = 72.67, SE = 3.46 for movement; M = 71.15, SE = 4.45 for non-movement), $F(1, 13) = 31.913$, $p < .001$, r = .84. No significant interaction effect was found between movement and expert level, $F(1, 13) = 0.007$, $p = .937$, r = .02.

For collapse we find that a significantly higher percentage of collapses occurred in the movement condition (M = 5.20, SE = 1.79) than in the non-movement condition (M = 3.31, SE = 1.09), $F(1, 13) = 7.805$, $p = .015$, r = .61. There was a significant effect of expert level, $F(1, 13) = 10.821$, $p = .006$, r = .67. In addition an interaction effect between movement and expert level was found, $F(1, 13) = 6.547$, $p = .024$, r = .58. Where the expert group revealed no difference in percentage of collapses between movement (M = 0.98, SE = 0.48) and non-movement (M = 0.81, SE = 0.30), the non-expert group have a higher percentage of collapses in the movement condition (M = 10.02, SE = 2.92) than in the non-movement condition (M = 6.16, SE = 1,82).

Joint agency and quality assessments: With respect to agency we find that participants significantly gave higher scores in the movement condition (M = 3.56, SE = 0.20) compared to the non-movement condition (M = 3.10, SE = 0.19), $F(1, 12) = 4.938$, p = .046, r = .54. There is no effect of expert level, $F(1, 12) = 3.252$, $p = .096$, r = .46, and no interaction effect between movement and expert level, $F(1, 12) = 0.002$, $p = .967$, r = .01.

The quality assessment in the movement condition is not significantly different from that in the non-movement condition, $F(1, 13) = 1.880$, $p = .194$, r = .36. There was a significant effect of expert level, indicating that experts gave higher annotation scores (M = 89.70, SE = 4.90 for movement; M = 83.99, SE = 5.90 for non-movement) than non-experts, (M = 69.03, SE = 5.56 for movement; M = 63.70, SE = 3.54 for non-movement) $F(1, 13) = 11.477$, $p = .005$, r = .68. No significant interaction effect was found between movement and expert level, $F(1, 13) = 0.002$, $p = .962$, r = .01.

Correlations of performance errors

All types of performance error are significantly correlated with one another. Table 1 shows the correlation values and their corresponding significance values for the three types of performance errors.

Table 1. Performance error correlations.

	Fluctuation vs. Narration		**Fluctuation vs. Collapse**		**Narration vs. Collapse**	
Trials	τ	p	τ	p	τ	p
All	-.67	< .001	.46	< .001	-.52	< .001
Movement	-.60	< .001	.50	< .001	-.57	< .001
No Movement	-.75	< .001	.46	< .001	-.50	< .001

Correlations of performance errors with agency and quality assessments

Fluctuation errors are negatively correlated with agency assessments, although correlation values are low (τ = -.15). The lower the fluctuation error, the higher the agency assessment value. Narration is positively correlated with quality assessments. The higher the percentage of predictable IOI classes, the higher the agency assessment value. Collapse errors are negatively correlated with quality assessments. The lower the percentage of collapses, the higher the agency assessment value

With respect to the quality assessments, higher correlations are found. Fluctuation is negatively correlated with quality assessment. The lower the fluctuation error, the higher the quality score. Narration is positively correlated with quality assessments. The higher the percentage of predictable IOI classes, the higher the quality assessment value. Collapse errors are negatively correlated with quality assessments. The lower the percentage of collapses, the higher the quality assessment value. Table 2 shows all Kendall's tau correlation coefficients and the corresponding significance values.

Table 2. Performance error correlations with subjective performance assessments.

		Fluctuation		Narration		Collapse	
	Trials	τ	***p***	τ	***p***	τ	***p***
Agency	All	-.15	.028	.18	.010	-.24	.001
	Movement	-.10	.311	.17	.083	-.31	.002
	No Movement	-.16	.126	.22	.030	-.26	.013
Quality	All	-.37	< .001	.40	< .001	-.41	< .001
	Movement	-.35	< .001	.35	< .001	-.48	< .001
	No Movement	-.35	< .001	.43	< .001	-.36	< .001

Discussion

The present paper investigated whether the quality of interaction, while performing music, plays a role in the establishment of joint action and group bonding experiences. The hypothesis that interaction quality plays a role was tested with singing dyads. We thereby focused on timing markers. We achieved three main outcomes. Firstly, we found correlations, albeit weak, between the sense of joint agency and measured fluctuation, narration and collapse errors. Contrary to our prediction, the highest degrees of joint agency reached by the dyads point to a SHARED rather than a WE sense of agency, particularly in the movement condition. Secondly, we found correlations between the self-annotated performance quality and measured fluctuation, narration and collapse errors. Although movement as such did not produce overall improvement in the quality of the performances, we observed a tendency for participants to reduce fluctuation errors while moving. On the contrary, non-expert dyads showed more collapse errors in that condition. These results point toward a different kind of effect of movement on micro- and macro-timing: when movement might possibly reduce micro-timing errors in general (recall that the difference was close to significant, with a medium effect size), it disrupts the performance on a macro-timing level for less experienced performers. Finally, we contributed to a novel and effective methodology and framework to analyse the objective quality of the interaction from the point of view of timing.

An important limitation of our study concerns the fact that the assessment of the feeling of joint agency was done after watching the performance recording, while moments of joint agency are supposed to occur during the performance. The correlation results are promising, but they point towards the need for a more refined method for estimating agency. The idea of using a hocket composition in our study was also inspired by Bolt et al. (2016)' s discovery that sequences of (twelve) tones played at the piano by pairs of non musicians, first by one subject and then by the second one, resulted in lower values of joint sense of agency compared to when the subjects alternated a tone after another. Moreover, these authors found that objective coordination between the subjects (measured by means of cross-correlation of the tones' onset series) impacted on joint agency, enhancing it when the coordination was strong. These findings are coherent with ours, though the data were obtained with different analytical tools and in a study that did not deal with expressive quality. In the present paper, we were also interested in the kind of joint agency such a performance could induce. Therefore we administered a questionnaire asking the subjects an assessment of their experienced sense of joint agency, stressing that the lowest values indicated an independent control, the medium values a shared control and the highest values a complete unity with the partner in controlling the musical joint action. Since the average collected values were in the medium range, our results point toward a SHARED rather than a WE sense of agency. This outcome complies, indeed, with Pacherie 's definition of the two kinds of joint agency (Pacherie, 2012), in particular when she suggests that a SHARED agency would ensue from a small group joint action, in which roles can be easily distinguishable. At this moment, we can speculate that this feature overcame the high similarity we intentionally established between the two scores. Indeed, according to Pacherie, the high predictability of, and, as a consequence, low necessity to keep oneself distinguishable from, the partner could have caused a WE-agency, rather than a SHARED agency (see Fairhurst et al. 2013 for a similar idea and some neuro-scientific possible account of it). Sticking to this result, we may then conclude that, on average, our musical task did not induce any boundary loss between the subjects in the pair, but we cannot exclude that the difficulty of the task contributed to prevent it. All in all, this finding adds to the debate on joint agency not only in musicology, but also in the wider domain of cognitive science (van der Wel et al. 2012, Dewey et al. 2014, Bolt & Loehr 2017).

Subjective self-annotations of the quality of the performance have to be treated carefully as well. The correlation results are promising and they seem to indicate that performance quality can be self-assessed in a proper way, although improvements to our slider approach in the video-stimulated recall protocol are still possible. Here, an important limitation of our study consisted in the latency between the recorded performance and the annotation the subjects did by means of the slider, meaning that the assessment cannot match perfectly the moments it refers to, but it is always a bit late. Furthermore, we asked the subjects to assess the quality of the performance as a whole, without focusing on timing, since we were interested also in other expressive features like pitch and tuning (whose analysis we are bracketing in the present study). Yet, given the crucial role of timing in music and its capacity to create social bonding in synchronization tasks (Hove & Risen, 2009, Wiltermuth & Heath 2009, Kokal et al. 2011), we assumed timing was the main feature to be analysed in our study. The good level of musicianship declared by our subjects, and visible in many of their performances, should bolster the validity of the correlation we found. Of course, not all couples reached the same quality levels, as it is manifest from both the objective and subjective measurements and from Figure 3, which shows the clustering of each couple's trials according to their fluctuation errors. Yet, we think that considering the relationships between those measurements gave us some hint about a proper treatment of the expressive quality in a singing dyad.

Also, the relatively large number of rejected data may induce some improvement of our paradigm. Indeed, most of the rejected trials were due to the fact that subjects did not comply with the experimental condition, either moving when they were supposed not to do so or singing completely differently than what was in the score. Some kind of feedback, either a visual or an auditory feedback, could inform the subject about his/her passing a given movement threshold, thus allowing to adjust for it. After all, both visual and auditory bio-feedback systems may be conceived of in order to adjust the performance itself according to the amount of (mainly fluctuation and narration) errors collected in a given time interval. This is how we see a relevant application of our method aiming at enhancing musical learning processes (see Moens & Leman 2015, for some applications of the same principle to running and walking to the music).

In this paper we developed also a novel and proper methodology to capture the interaction of a singing dyad. While the method was applied to the emergent timing of both singers, the method allows an analysis of each singer separately, despite the non-stationarity of the data. In accordance with the recent emphasis on the predictive brain theoretical framework (Friston 2010, Clark 2016), also in music studies (Vuust et al. 2009, Koelsh et al. 2019), we applied a Bayesian inference approach to dynamically analyse a semi-hocket interaction between two subjects. In fact, a singing dyad can be conceived of as a dynamical system whose components constrain each other's unfolding performance (Konvalinka & Roepstorff 2012, Müller & Lindenberger 2011, Müller et al. 2018), considering its variability and correcting for it, when needed. A sequential Bayesian process allowed for a non-stationary analysis in the form of a continuous updating of timing-error minimisation. We focused on timing and identified fluctuation, narration and collapse errors as objective, third-person markers of the quality of a musical interaction, exploiting the idea that the "superordinate system", i.e. the dyad, rather than the single singer, constructed predictions of latent variables that keep track of the timing of each relative inter-onset-interval. This approach has the advantage that we look finer in time than a method that would focus on the overall tempo. Obviously, it can be questioned whether this construct has any psychological plausibility, yet the emergence of latent variables is a known phenomenon, and in full agreement with the predictive coding approach. For example, the concept of latent variables that work as predictors for observable/measurable action can be compared with the two processes postulated to correct errors in a sensorimotor synchronization task at the individual level, phase correction and period correction, the former being an almost automatic process with which fluctuation errors can be equated, the latter requiring a conscious effort comparable to the one needed to overcome narration errors (Wing et al. 2010, Repp & Su, 2013). The distinction between fluctuation, narration and collapse errors was introduced in order to deal with typical performance errors. Fluctuation may be related to subconscious active sampling in order to be able to update the latent variable on timing. Further research is needed to refine its sources of variability. Narration relates to a symbol-based account of the performance and therefore, we assume that it has a cognitive origin related to memory and sequencing. While collapse errors induce a complete breakdown of the performance, the singers may still cope with narration errors (possibly with period correction), even if they surely threaten the quality of the performance. We believe that the Bayesian inference framework offers a useful method for accessing musical expression in high quality music performance. As our concept is based on relative IOIs, the method offers the perspective that it can be applied to groups comprising three and more singers and musicians.

Finally, movement did not improve the performance timing, but the fact that the worse couples made more collapse errors in the movement condition, along with the higher joint agency values reported in that condition and a tendency for all participants to reduce their fluctuation errors in that condition, suggests that above a certain level movement may impact on the overall quality of the performance. In particular, this result could imply that, while for bad couples movement constitutes an interference with their task, good couples may benefit from it at a micro-timing level. This hypothesis is compatible with a Bayesian approach insofar as bad couples, by definition, find it difficult to both coordinate their movements with the music and their singing with the partner's, that is, predicting the music and the partner at the same time. On the other hand, active inference may be enhanced by moving for those couples that are already fluent, but can take further advantage from moving at a micro-timing level. However, further research is surely needed to better disentangle the network of dynamic processes that is constituted by prediction, agency and movement in musical expressive moments (Leman, 2016).

As far as we know, this is the first study that applies principles of the predictive coding approach to a social musical interaction. And it does so by stressing the dynamic character of the interaction thanks to a parameter, the relative IOI, which treats two subjects as one, hence taking seriously the Gestalt concept that the whole is more than the sum of its parts. The same idea is implicit in the concept of participatory sense-making (De Jaegher & Di Paolo 2007), which emphasizes that the sense of a joint action is not given in advance, but it is co-constituted by the interactive subjects. In a musical context, thereby, the musical object is not constituted either by the score or by the representations in the minds of each musician, not even by the auditory event in itself, but rather by the embodied interaction of the musicians on the fly (Schiavio & De Jaegher 2017). The focus on the interaction, rather than on the single components of it, increases the complexity of studying an already complex phenomenon like music, although also in the domain of cognitive neurosciences several appeals have been recently made toward such a perspective change. For example, Schillbach et al. (2013) write that "After more than a decade of research, the neural mechanisms underlying social interaction have remained elusive and could – paradoxically – be seen as representing the "dark matter" of social neuroscience" (ibidem: 394). Hyper-scanning, the simultaneous acquisition of cerebral data from two or more subjects, is a promising technique to approximate such ambitious aim (Kovalinka & Roepstorff 2012, Babiloni & Astolfi 2014). Indeed, though not yet analysed, not only did our experiment carry out a motion capture collection of data from the singing dyads, but it also planned the physiological recording of skin conductance by means of portable bracelets. Moreover, we are working exactly on the possibility to simultaneously electroencephalography (EEG) recording two interacting musicians, in search of the brain basis of social embodied music interaction. Such an empowered set-up would likely allow both to test the psychological plausibility of a dynamic marker of timing as the one we devised in the present paper and to identify possible dynamic neural markers of timing and other musical features and processes (see also Osaka et al. 2015, Nozaradan et al. 2016, Pan et al. 2018). Ultimately, such enterprise would probably require a thorough theoretical synthesis between embodied and predictive approaches to (music) cognition, of which the present work can be seen as a first empirical application.

References

Adams R., Stewart S. & Friston K. (2013), Predictions not commands. Active inference in the motor system. *Brain Structure and Function*, 218:611–643

Arbib, M. (Ed.) (2013). *Language, music, and the brain: A mysterious relationship*. Cambridge, MA: MIT Press.

Bolt, N. K., Poncelet, E. M., Schultz, B. G., & Loehr, J. D. (2016). Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Consciousness and Cognition*, 46, 173–187.

Bolt, N. & Loehr, J., (2017). The predictability of a partner's actions modulates the sense of joint agency. *Cognition* 161, 60–65.

Chanda, M. L. & D. J. Levitin (2013). The neurochemistry of music. *Trends in cognitive sciences. 17*(4), 179–193.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Clark A. (2016). *Surfing Uncertainty*. Oxford University Press. Oxford

D'Ausilio, A., Novembre, G., Fadiga, L. & Keller, P.E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, *19*(3), 111-114.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences, 6*(4), 485-507.

Dewey, J. A., Pacherie, E. & Knoblich, G. (2014). The phenomenology of controlling a moving object with another person. *Cognition*, 132, 383–397.

Eerola, T., Jakubowski, K., Moran, N., Keller, P. & Clayton, M. (2018). Shared Periodic Performer Movements Coordinate Interactions in Duo Improvisations. *Royal Society Open Science* 5(2): 171520.

Fairhurst, M.T., Janata, P., Keller, P.E., (2013). Being and feeling in sync with an adaptive virtual partner: brain mechanisms underlying dynamic cooperativity. *Cerebral Cortex* 23, 2592–2600.

Freeman, W.J., 2000. A neurobiological role of music in social bonding. In: Wallin, N.L., Merker, B., Brown, S. (Eds.), *The Origins of Music*. Cambridge, MA: The MIT Press, pp. 411–424

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Fritz, T., S. Hardikar, M. Demoucron, M. Niessen, M. Demey, O. Giot, Y. Li, J.-D. Haynes, A. Villringer & Leman M. (2013). Musical agency reduces perceived exertion during strenuous physical performance. *PNAS 110*(44), 17784–17789.

Gallagher S. (2005). *How the Body shapes the Mind*. Oxford University Press

Gallagher S. & Allen M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese.*195:2627–2648

Goebl, W. & Palmer, C., (2009). Synchronization of timing and motion among performing musicians. *Music Perception.* 26, 427–438.

Haggard P. & Eitam B. (2015) (Eds.), *The Sense of Agency*. Oxford, England: Oxford University Press

Honing, H., Cate, C., Peretz, I., Trehub, S.E., (2015). Without it no music: cognition, biology and evolution of musicality. *Philosophical Transaction of the Royal Society,* B 370.

Hove, M. J. & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27, 949–960.

Keller P. E. (2014). Ensemble performance: interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in Music Performance: Empirical Approaches Across Styles and Cultures* (pp. 260–282), Oxford, UK: Oxford University Press.

Keller, P. E., Novembre, G., & Loehr, J. D. (2016). Musical ensemble performance: Representing self, other, and joint action outcomes. In S. Obhi & E. Cross (Eds.), *Shared representations: Sensorimotor foundations of social life* (pp. 280–311). Cambridge, UK: Cambridge University Press.

Keller PE, Novembre G, Hove MJ. (2014) Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transaction of the Royal Society* B, 369: 20130394

Keller, P.E., Knoblich, G., Repp, B.H., (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition* 16, 102–111.

Koelsch S. Vuust P. Friston K. (2019) Predictive Processes and the Peculiar Case of Music. *Trends in cognitive sciences.* Vol. 23, n. 1, 63-77

Kokal I, Engel A, Kirschner S, Keysers C (2011). Synchronized Drumming Enhances Activity in the Caudate and Facilitates Prosocial Commitment - If the Rhythm Comes Easily. *PLoS One* 6(11)

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, 6.

Kreutz G. (2014). Does singing facilitate social bonding? *Music and Medicine*, v. 6 (2).

Iyer, V. (2002). Embodied mind, situated cognition, and expressive micro-timing in African-American music. *Music Perception*, 19, 387–414.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press

Leman M. (2016). *The Expressive Moment*. MIT Press. Cambridge (MA, USA)

Lesaffre M., Maes P.J., Leman M. (eds.) (2017), *Routledge Companion to Embodied Music Interaction*. New York and London: Routledge

Loehr, J. D., Large, E. W., & Palmer, C. (2011). Temporal coordination and adaptation to rate change in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 37*(4), 1292–1309.

Loehr, J. D., Kourtis D., Vesper C., Sebanz N. & Günther K. (2013). “Monitoring Individual and Joint Action Outcomes in Duet Music Performance.” *Journal of cognitive neuroscience* 25(7):1049–61.

MacDonald R., Kreutz G. & Mitchell L., (ed.) (2012) *Music, Health and Wellbeing*. Oxford: Oxford University Press

Maes, P-J., Leman, M., Palmer, C., & Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in Psychology, 4*(1008)

Malloch S. & Trevarthen C. (2009). *Communicative Musicality*. Oxford University Press, UK.

McNeill, W.H., (1995). *Keeping together in time*. Harvard University Press, Cambridge, MA.

Moens B. & Leman M. (2015). Alignment strategies for the entrainment of music and movement rhythms. *Annals of the New York Academy of Science.* 1337, 86–93.

Müller, V. & U. Lindenberger. (2011). Cardiac and respiratory patterns synchronize between persons during choir singing. *PLoS One* **6:** e24893.

Müller V., Delius JAM., Lindenberger U. (2018). Complex networks emerging during choir singing. *Annals of the New York Academy of Sciences*. 1431(1):85-101.

Nozaradan S, Peretz I, Keller P. (2016). Individual differences in rhythmic cortical entrainment are correlated with predictive behaviour in sensorimotor synchronization. *Nature Scientific Reports*, 6, 20612

Overy, K. & Molnar-Szakacs, I., (2009). Being together in time: musical experience and the mirror neuron system. *Music Perception*. 26, 489–504

Pacherie, E., (2012). The phenomenology of joint action: self-agency vs. joint-agency. In: Seemann, A. (Ed.), *Joint Attention: New Developments*. MIT Press, Cambridge, MA, pp. 343–389.

Repp, B.H., Su, Y.H., 2013. Sensorimotor synchronization: a review of recent research (2006– 2012). *Psychonomic Bulletin Review*. 20, 403–452.

Rohrmeier, M.A. & Koelsch, S. (2012) Predictive information processing in music cognition. A critical review. *Int. J. Psychophysiology* 83, 164–175

Salimpoor, V.N., Zald D., Zatorre R., Dagher A. & MacIntosh A.R. (2015) Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences.* 19, 86–91

Schiavio & De Jaegher (2017) Participatory sense-making in joint musical practice. In (. Lesaffre M., Maes P.J., Leman M.), 31-39.

Sebanz, N., Bekkering, H., Knoblich, G., (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences.* 10, 70–76.

Thompson E. & Varela (2001). Radical Embodiment: neural dynamics and consciousness. *Trends in cognitive sciences* Vol.5 No.10, 418-425

Vesper C, Butterfill S, Knoblich G. & Sebanz N. (2010) A minimal architecture for joint action. *Neural Networks.* 23, 998 – 1003

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C. & Roepstorff, A. (2009). Predictive coding of music– brain responses to rhythmic incongruity. *Cortex* 45, 80-92

Walton, A., Richardson, M. J., Langland-Hassan, P. & Chemero, A. (2015). Improvisation and the self- organization of multiple musical bodies. *Frontiers in Psychology*. 6, 313.

Wan C., Rüber T., Hohmann A. and Schlaug G. (2010) The Therapeutic Effects of Singing in Neurological Disorders. *Music Perception.* April 1; 27(4): 287–295.

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2012). The sense of agency during skill learning in individuals and dyads. *Consciousness and Cognition*, 21, 1267–1279.

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

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

Wing, A. M., Doumas, M., & Welchman, A. E. (2010). Combining multisensory temporal information for movement synchronisation. *Experimental Brain Research*, 200, 277–282.

Chapter 6. Conclusions

We begin the present chapter trying to answer the different open questions we posed in the Introduction (§ 1.7.1).

1) Several studies in the last two decades have demonstrated that also non-musicians are endowed with some proto-musical competences, both in perception and production (Koelsch et al. 2000, Mehr et al. 2019), but only few of them have focused on joint musical actions. By means of alternate joint tapping we discovered that also pairs of non musicians are able to adapt their timing the their partner's, a competence that was held to be a musicians' exclusive so far (Keller 2008). We did it by computing their asynchronies with respect to the metronome and cross-correlating them, finding that subjects tended to imitate their partner's asynchronies, rather than correcting for them, whether they occurred sooner or later than the metronome beats. Working as part of an embodied language, the rhythmic component provided by the metronome mutually entrained the basic motor actions of the interacting dyad, before any conscious awareness of the process from the subjects' side.
2) Taking advantage of a well-known illusion of ownership (the rubber hand illusion), we found that mutual adaptive timing was established even in a condition in which the partner's hand is embodied. In other words, the phenomenon of entrainment emerged not only when the subjects faced each other, but also in an apparently non-social condition like the "egocentric" condition, overcoming the embodiment of the partner's hand. We interpreted such a finding as corroborating the idea of a universal nature of musicality as an embodied language, due to the intrinsic characteristics of the joint proto-musical system. However, the physiological results collected thanks to single-pulse TMS enhance the conclusion that the "allocentric" condition was judged as the social one, resulting in higher cortico-spinal excitability than both the solo and the egocentric conditions (in keeping with the mirror neurons literature, Fadiga et al. 1995, Novembre et al. 2012). On the contrary, the embodiment of the partner led the subject to cope with the latter condition as if no partner was present at all,

resulting in a cortico-spinal activation comparable to the solo (and the baseline) condition.

3) Previous literature has already shown the pro-social effects of synchronizing with a partner (Wiltermuth & Heath 2009, Kokal et al. 2011), but, to our knowledge, no one explored the possibility that a musical interaction might remap the peri-personal space of two musicians according to the nature of such interaction, be it cooperative or uncooperative. It turned out that, after a jazz performance with an uncooperative partner, who played the wrong harmonic sequence of a jazz tune, a musician's peri-personal space shrank, as if a musical joint action became impossible. We interpreted this result as evidence that, insofar as music and musicality are intrinsically social (embodied languages), a musical interaction has a measurable impact on the perception of the space between two (or more) subjects. Consistently with an "extended mind" hypothesis (Clark & Chalmers 1998), as tools can be incorporated into their users' body-schema and as body parts of another person can be embodied (see above), we speculate that the peri-personal spaces of two (or more) interacting subjects may merge, as portrayed by the concept of "mutual incorporation" (Fuchs & De Jaeger 2009).
4) The audio-tactile integration task we exploited in order to measure peri-personal space (Serino et al. 2007) allowed us to compare our sample of jazz musicians with a sample of non-musicians in the experimental baseline condition. As expected, the former group exhibited higher multisensory integration competences, arguably due to the musicians' sensorimotor training with their instrument and (to a lesser extent) singing, that brings about well-known cortical-subcortical reorganizations (Munte et al. 2002, Zimmermann & Lahav 2011). Moreover, it is likely that musicians exploit a sensorimotor mirror network both when listening to musical excerpts (Herholz & Zatorre 2012) and when observing music-related movements (Candidi et al. 2014). This is why in a further experiment we tested a sample of wind musicians, under the hypothesis that their cortico-spinal excitability in FDI could reveal their expertise in integrating (tactile) holding of a trumpet with the sound of a trumpet. Given the small sample, our results have to be considered as temporary, showing a trend toward a lower cortico-spinal excitability in musicians, compared to non-

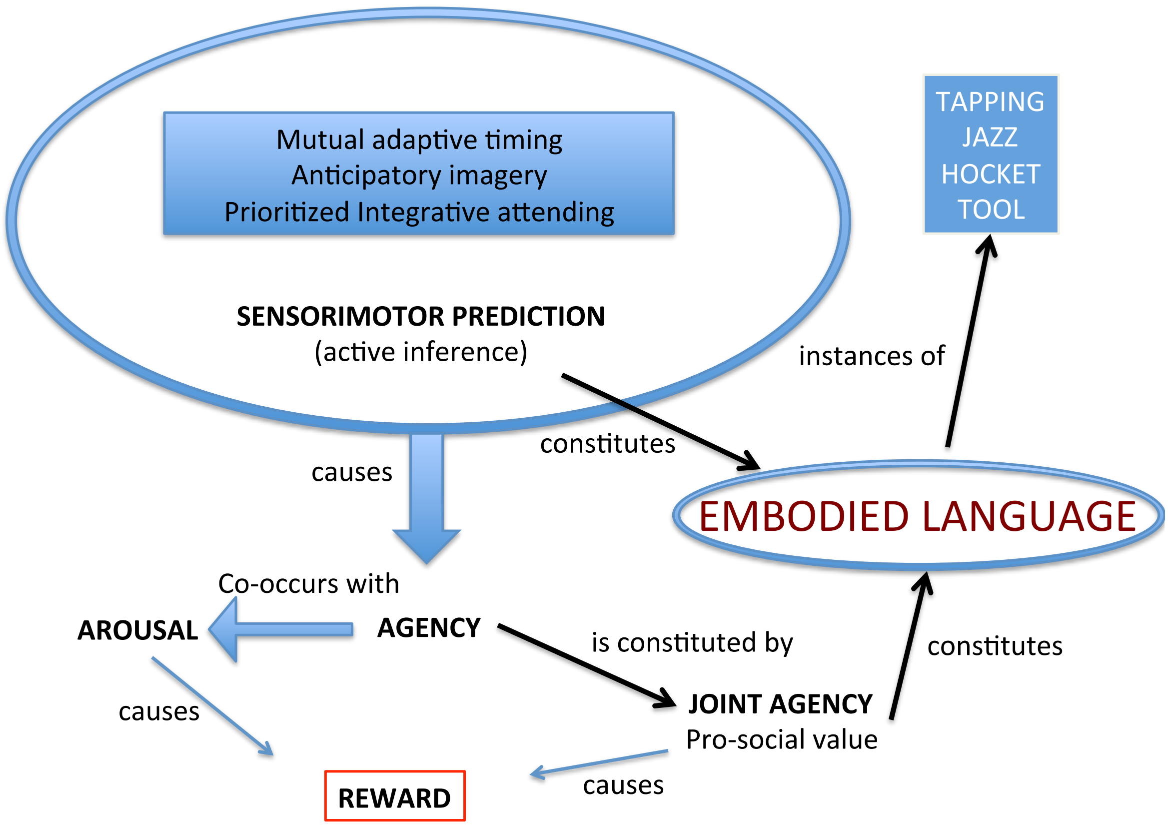
musicians, while listening to white noise, compared to trumpet tones, independently of the tool held. The lower sensitivity to white noise exhibited by musicians may be taken as a marker of their sensorimotor expertise, but further testing is needed to draw stronger conclusions about both populations' multisensory competences.

5) After the short deviation about multisensory integration, we came back to embodied interaction thanks to a hocket singing paradigm in which pairs of musicians' expressive quality was tested exploiting a Bayesian analysis method. Such a method allowed us to explore timing without assuming stationarity in the performance (contrary to the majority of the present studies, e.g. Goebl & Palmer 2009, Clayton et al. 2019), under the hypothesis that predictive models carry out a continuous updating of timing-error minimisation, comparing the actual inter-onset intervals between the two singers' tones with the predicted ones. Indeed, a singing dyad can be conceived of as a dynamical Gestalt whose components constrain each other's unfolding performance through the medium of that embodied language represented by music (Walton et al. 2015, Müller et al. 2018). Interestingly, the movement condition showed a different effect according to the expertise, in that, while all the pairs tended to be more correct in the micro-timing (what we called, "fluctuation" errors), lower quality pairs disrupted significantly more often the performance in the movement, compared to the non-movement, condition (they made more "collapse" errors).

6) On average, the singers reported a feeling of SHARED, rather than WE, agency, meaning that they felt to control only part of the joint action, rather than feeling as a single entity with their partner while accomplishing that action (Pacherie 2012). However, it turned out that this subjective parameter was correlated with the objective parameter devised by our Bayesian analysis, i.e. duration errors (fluctuation, narration and collapse errors), showing its plausibility as a (subjective) marker of the expressive interaction. Even more highly correlated than joint agency with the timing markers of the interaction quality were the self-annotations of that very quality made by the musicians. Our results suggest that the small size of the group (a pair) has prevented the musicians to merge into one (WE-agency), although the score was built to be highly symmetrical and

predictable (Bolt et al. 2016, Bolt & Loehr 2017), thus trying to induce such a merging.

The concept of music and musicality as an embodied language around which the present work revolves can now be summarised by means of a model. As we hinted at more than once in the Introduction, the ambition of such a model would be to integrate predictive coding and embodied frameworks, borrowing from two already existing accounts of musical interaction, that is, Keller's (2008) and Leman's (2016).



The figure above intends the three sensorimotor competences necessary to play music together, according to Keller, as components of a prediction device, which makes use of active inference during a musical interaction, be it an individual or a joint action, be it simply listening or producing music. The rewarding effect of interacting with music in such a variety of ways, stressed by Leman, may be reduced to one crucial factor, that is, agency, as a consequence of embodied sensorimotor predictions. While arousal is an agency ensuing feature that we have not developed here (but see some remarkable studies like Fritz et al. 2013 or Tarr et al. 2014), joint agency is how I see agency in a musical context. Indeed, given the pro-social value of musicality (and, then, of music) as an embodied language, agency in such contexts is not simply the feeling of being in control of a given individual action, as in grasping an object for an ordinary action, but it has to be characterized by a "joint" component, implying the more or less evident presence of one or more musically interacting subjects. The weak version of this model is easily applied to a real ensemble performance as in the dyadic interactions we explored in the present thesis, but a stronger version would identify a social component also in individual interactions with music, as in the case of our experiment with trumpets as musicians' (embodied) tools.

According to our results and the previous model, next studies might investigate the following issues, e.g.:

1) As we saw, one advantage of tapping studies is the possibility to test non-musicians, but will a dynamic method as the one we used in Chapter 5 confirm the results we obtained in Chapter 2 with cross-correlations?
2) Would a paradigm as the one we used in Chapter 2 be effective from a clinical point of view, for example, to recover the functionality of an arm after a stroke that injured it? In particular, the egocentric condition should be compared with the allocentric one.
3) In Chapter 3 we employed an indirect measure of peripersonal space. It would be interesting to compare it with different ways of measuring it, like verbal reports, and try to differentiate between the peripersonal and the "interpersonal" space (as other studies did in non-musical contexts).
4) To what extent may the expressive quality of a musical interaction (see Chapter 5) impact on the perception of the peripersonal space?
5) Can the Bayesian method employed in Chapter 5 be extended to interactions between more than two musicians, e.g. to combos or orchestras?
6) Which parameters may induce a WE, rather than a SHARED agency, during/after a musical interaction?
7) Correlating hyper-scanning measures with joint sense of agency, according to the musical genre and the quality of a musical interaction.

8) Would Orbicularis Oris or tongue muscles show the cortico-spinal excitability predicted by Chapter 4 in wind musicians while integrating the touch and the sound of a wind instrument?
9) Would cortico-spinal activation increase in a multisensory compared to a one sensory modality musical condition? From this point of view, it would be worthwhile comparing an observation to an interaction condition.

To sum up, in this thesis I tried to explore several aspects of making music in dyads, from a very basic proto-musical action, like tapping, to more sophisticated contexts, like playing a jazz standard and singing a hocket melody, passing through that crucial junction represented by the use of the musical instrument (a trumpet, in our case). If social interaction is the default mode by which humans communicate with their environment (Hari et al. 2015), music and musicality conceived of as an embodied language may arguably provide a route toward its navigation. The above questions are only a tiny part of the questions worth posing in a research program devoted to better understand the brain and behavioural bases of "musicking" together (Small 1998). Given the complex nature of the musical phenomenon, encompassing biological and cultural aspects, it will not be surprising to see interdisciplinary efforts in the near future, that put together evolutionary biologists, neuroscientists, psychologists, musicologists, philosophers as well as musicians. The present work aims to be but a drop in this sea, whose boundaries remain unexplored.

References

To avoid repetitions, references of the Introduction are included here, but references of § 1.7.2 may be found after each chapter.

Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., & Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience*, *19*(9), 2609-2612.

Babiloni F. & Astolfi L. (2014) Social neuroscience and hyperscanning techniques: Past, present and future, *Neuroscience and Biobehavioral Reviews* 44, 76–93

Babiloni, C., Buffo, P., Vecchio, F., Marzano, N., Del Percio, C., Spada, D., Rossi, S., Bruni, I., Rossini, P.M. & Perani D., (2012). Brains in concert: Frontal oscillatory alpha rhythms and empathy in professional musicians. *NeuroImage* 60 (1), 105–116

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unravelling multisensory integration: patchy organization within human STS multisensory cortex. *Nature neuroscience*, *7*(11), 1190.

Becchio C., Koul A., Ansuini C., Bertone C. & Cavallo A. (2018) Seeing mental states: An experimental strategy for measuring the observability of other minds *Physics of Life Reviews 24, 67–80*

Berti, A., & Frassinetti, F. (2000). When far becomes near: Remapping of space by tool use. *Journal of Cognitive Neuroscience*, 12, 415–420

Berti A. & Pia L. (2006) Understanding Motor Awareness Through Normal and Pathological Behavior. *Current Directions in Psychological Science* 15 (5), 245-25

Bolt, N. K., Poncelet, E. M., Schultz, B. G., & Loehr, J. D. (2016). Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Consciousness and Cognition*, 46, 173–187

Bolt, N. & Loehr, J. (2017). The predictability of a partner's actions modulates the sense of joint agency. *Cognition* 161, 60–65.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, *391* (February), 756.

Brattico E. & Vuust P. (2017) Brain-to-brain coupling and culture as prerequisites for musical interaction. In *Routledge Companion to Embodied Music Interaction* (eds. Lesaffre M., Maes P.J., Leman M.), 249-257. New York and London: Routledge

Caccioppo J., Bernston G. & Decety J. (2010) Social neuroscience and its relationship to social psychology**,** *Social Cognition*, Vol. 28, No. 6, 675–685

Candidi, M., Maria Sacheli, L., Mega, I., & Aglioti, S. M. (2012). Somatotopic mapping of piano fingering errors in sensorimotor experts: TMS studies in pianists and visually trained musically naives. *Cerebral cortex*, *24*(2), 435-443.

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLoS One*, 7, e44306

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press

Clark, A., & Chalmers, D. (1998). The extended mind. *Analyses, 58*, pp. 7-19.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge

Clark, A. (2016). *Surfing uncertainty*. Oxford: Oxford University Press

Clarke EF. (2005) *Ways of Listening. An Ecological Approach to the Perception of Musical Meaning*. New York: Oxford University Press;

Clarke EF., De Nora T. & Vuoskoski J. (2015) Music, empathy and cultural understanding, *Physics of Life Reviews* 15, 61-88

Clayton M. (2012). What is entrainment? Definition and applications in musical research. *Empirical Musicology Review*. 7 (1–2), 49–56.

Clayton, M. Jakubowski, K. & Eerola, T. (2019) 'Interpersonal entrainment in Indian instrumental music performance: synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure.', *Musicae scientiae*. 23(3): 304-331

Colling, L. & Williamson, K. (2014). Entrainment and motor emulation approaches to joint action: alternatives or complementary approaches? *Frontiers in Human Neuroscience*. 8 (754)

Cross I. (2014). Music and communication in music psychology. *Psychology of Music* 42 (6), 809–819.

D'Ausilio A, Novembre G, Fadiga L. & Keller PE. 2015 What can music tell us about social interaction? *Trends in Cognitive Science*. 19, 111 – 114.

De Jaegher H. & Di Paolo E. (2007) Participatory sense- making an enactive approach to social cognition. *Phenomenology and the Cognitive Science* 6, 485 – 507.

Doffman M.R., (2008). *Feeling the Groove: Shared Time and Its Meanings for Three Jazz Trios* (Ph.D. Dissertation, Music Department). Open University

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology*, *73*(6), 2608-2611.

Freeman W.J. (2000). A neurobiological role of music in social bonding. In: Wallin, N.L., Merker, B., Brown, S. (Eds.), *The Origins of Music*. Cambridge, MA: The MIT Press, pp. 411–424.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, *10*(86), 20130475.

Friston, K., & Frith, C. D. (2015). A duet for one. *Consciousness and Cognition*, *36*, 390–405.

Fritz, T., S. Hardikar, M. Demoucron, M. Niessen, M. Demey, O. Giot, Y. Li, J.-D. Haynes, A. Villringer & Leman M. (2013). Musical agency reduces perceived exertion during strenuous physical performance. *PNAS 110*(44), 17784–17789.

Fuchs, T., & De Jaegher, H. (2009). Enactive Intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8*(4), 465-486

Gallagher S. & Allen M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese.*195:2627–2648

Gallagher S. (2017). *Enactive Interventions*. Oxford: Oxford University Press

Goebl, W. & Palmer, C., (2009). Synchronization of timing and motion among performing musicians. *Music Perception.* 26, 427–438

Gordon C., Iacoboni M. & Balasubramaniam R. (2018) Multimodal music perception engages motor prediction: a TMS study, *Frontiers in Neuroscience.* 12:736

Graziano, M. S. A. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences*, *96*(18), 10418–10421.

Graziano, M. S. A. & Gandhi, S. (2000). Location of the polysensory zone in the precentral gyrus of anesthetized monkeys. *Experimental Brain Research*, 135, 259–266.

Haggard P. & Eitam B. (2015) (Eds.), *The Sense of Agency*. Oxford, England: Oxford University Press

Hari R., Henriksson L., Malinen S. & Parkkonen L. (2015) Centrality of Social Interaction in Human Brain Function, *Neuron* 88, 181-193

Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A.O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience, 17*(2), pp. 282-293.

Herholz, S. C. & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behaviour, function and structure. *Neuron,* 76, pp. 486–502.

Honing H, ten Cate C, Peretz I. & Trehub SE. (2015) Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society*. B 370: 20140088.

Hurley, S. (1998). *Consciousness in Action*. Harvard University Press, London

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, 7, 2325–2330

Iyer, V. (2002). Embodied mind, situated cognition, and expressive micro-timing in African-American music. *Music Perception*, 19, 387–414

Janata, P., Tomic, S. T. & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol. Gen.* 141, 54–75.

Keller PE. (2008) Joint action in music performance. In *Enacting intersubjectivity: a cognitive and social perspective to the study of interactions* (eds F. Morganti, A. Carassa, G. Riva), pp. 205 – 221. Amsterdam, The Netherlands: IOS Press.

Keller P.E. & Appel M. (2010) Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*. 28, 27–46.

Keller PE, Novembre G. & Hove MJ. (2014) Rhythm in joint action: psychological and neuro-physiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society* B 369: 20130394.

Kelso, J.A. (1995). *Dynamic Patterns*. MIT Press, Cambridge MA.

Koelsch, S., Gunter, T., Friederici, A. & Schroeger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience*. 12/3, 520–541.

Koelsch S. Vuust P. & Friston K. (2019) Predictive Processes and the Peculiar Case of Music. *Trends in cognitive sciences.* Vol. 23, n. 1, 63-77

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, *297* (5582), 846-848.

Kokal I, Engel A, Kirschner S & Keysers C (2011). Synchronized Drumming Enhances Activity in the Caudate and Facilitates Prosocial Commitment - If the Rhythm Comes Easily. *PLoS One* 6(11)

Konvalinka I. & Roepstorff A. (2012) The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in Neuroscience* Vol. 6:215

Leman, M. (2007). *Embodied Music Cognition*. MIT Press, Cambridge MA.

Leman, M. (2016). *The Expressive Moment*. MIT Press, Cambridge MA.

Lerdahl, F., & Jackendoff, R (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

London J. (2012) *Hearing in Time: psychological aspects of musical meter*. New York, NY: Oxford University Press.

Malloch S. & Trevarthen C. (2009). *Communicative Musicality*. Oxford University Press, UK.

Maravita, A. & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, 8, 79–86

Marsh K., Richardson M. & Schmidt R. (2009) Social Connection Through Joint Action and Interpersonal Coordination, *Topics in Cognitive Science* 1, 320–339

McNeill, W.H. (1995). *Keeping together in time*. Harvard University Press, Cambridge, MA.

Mehr S. A. et al. (2019) Universality and diversity in human song, *Science,* vol. 366, 6468

Menary, R. (Ed.) (2010), *The Extended Mind*. MIT Press, Cambridge, MA.

Mithen S. (2005) *The Singing Neanderthals: The Origins of Music, Language, Mind and Body.* London: Weidenfeld and Nicolson

Moens B. & Leman M. (2015). Alignment strategies for the entrainment of music and movement rhythms. *Annals of the New York Academy of Science.* 1337, 86–93.

Müller V., Delius JAM., Lindenberger U. (2018). Complex networks emerging during choir singing. *Annals of the New York Academy of Sciences*. 1431(1): 85-101.ù

Munte, T., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3, 473–478.

Nettl, B. (2005). *The Study of Ethnomusicology: 31 issues and concepts*. University of Illinois Press, Urbana-Chicago

Noë, A. (2004). *Action in Perception*. MIT Press, Cambridge.

Newen, L. De Bruin, & S. Gallagher (Eds.) (2018), *The Oxford Handbook of 4E Cognition.* Oxford, UK: Oxford University Press

Newman-Norlund RD, van Schie HT, van Zuijlen AM. & Bekkering H. (2007) The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience* 10, 817 – 818.

Novembre G, Ticini LF, Schutz-Bosbach S. & Keller PE. (2012) Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex* 22, 2894 – 2903.

Novembre G, Ticini LF, Schu¨tz-Bosbach S. & Keller PE. (2014) Motor simulation and the coordination of self and other in real-time joint action. *Social Cognitive Affective Neuroscience* 9, 1062 – 1068.

Osaka N., Minamoto T., Yaoi K., Azuma M., Minamoto Shimada Y. & Osaka M (2015) How Two Brains Make One Synchronized Mind in the Inferior Frontal Cortex: fNIRS-Based Hyper-scanning During Cooperative Singing. *Frontiers in Psychology*. 6:1811

Overy, K. & Molnar-Szakacs, I., (2009). Being together in time: musical experience and the mirror neuron system. *Music Perception*. 26, 489–504.

Pacherie, E., 2012. The phenomenology of joint action: self-agency vs. joint-agency. In: Seemann, A. (Ed.), *Joint Attention: New Developments*. MIT Press, Cambridge, MA, pp. 343–389.

Pan Y., Novembre G., Song B., Li X. & Hu Y. (2018) Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *NeuroImage* 183, 280–290

Phillips-Silver, J. & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science* 308, 1430.

Phillips-Silver J. & Keller PE. (2012) Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*. 6, 26.

Phillips-Silver, J., Aktipis, C.A. & Bryant, G.A. (2010). The ecology of entrainment: foundations of coordinated rhythmic movement. *Music Perception*. 28, 3–14.

Reddy, V. & Uithol, S. (2015), Engagement: Looking beyond the mirror to understand action understanding. *British Journal of Developmental Psychology*.

Repp B.H. & Su Y.H. (2013) Sensorimotor synchronization: a review of recent research (2006– 2012). *Psychonomic Bulletin Review*. 20, 403 – 452.

Rizzolatti, G., Scandolara, C., Matelli, M. & Gentilucci, M., (1981). Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behavioural and Brain Research.* 2 (2), 147–163

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain. How our Minds Share Actions and Emotions*. Oxford University Press, Oxford.

Rizzolatti G. & Sinigaglia C. (2010) The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Review Neuroscience*. 11, 264 – 274.

Salimpoor, V.N. et al. (2015) Predictions and the brain: how musical sounds become rewarding. *Trends in Cognitive Sciencse*. 19, 86–91

Schiavio, A., & Menin, D. (2013). Embodied music cognition and the mediation technology. A critical review. *Psychology of Music, 41*(6), pp. 804-814.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioural & Brain Sciences*. *36*, 393–414.

Sebanz, N., Bekkering, H. & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences.* 10, 70–76.

Serino A., Bassolino, M., Farnè, A., & Làdavas, E. (2007). Extended multisensory space in blind cane users. *Psychological Science*, *18*(7), 642–648.

Small C. (1998) *Musicking: The Meanings of Performing and Listening*, Wesleyan University Press

Soliman, T. M., Ferguson, R., Dexheimer, M. S. & Glenberg, A. M. (2015) Consequences of joint action: Entanglement with your partner. *Journal of Experimental Psychology: General* 144(4): 873-888.

Stephan, M. A., Lega C., & Penhune, V. B. (2018). Auditory prediction cues motor preparation in the absence of movements. *NeuroImage*, *174*, 288-296.

Su, Y.-H., & Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological Research* 76, 373–382

Tarr B, Launay J. & Dunbar RIM (2014). Music and social bonding: “self-other” merging and neurohormonal mechanisms. *Frontiers in Psychology*. Vol. 5, 1096

Teneggi, C., Canzoneri, E., di Pellegrino, G. & Serino, A. (2013). Social modulation of peripersonal space boundaries. *Current Biology*, 23(5), 406–411.

Thompson E., & Varela, F. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*, pp. 418-425.

Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception, 17,* 115–126.

van Noorden, L. & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research.* 28, 43–66.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. MIT Press, Cambridge, MA.

Vesper C, Butterfill S, Knoblich G. & Sebanz N. (2010) A minimal architecture for joint action. *Neural Networks.* 23, 998 – 1003

Volpe G, D’Ausilio A, Badino L, Camurri A. & Fadiga L. (2016) Measuring social interaction in music ensembles. *Philosophical Transactions of the Royal Society*. B 371: 20150377

Walton, A., Richardson, M. J., Langland-Hassan, P. & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*. 6, 313

Walton A., Washburn A., Langland-Hassan, P. & Chemero, A., Kloos H., Richardson, M. J. (2018) Creating Time: Social Collaboration in Music Improvisation. *Topics in Cognitive Science* 10(1): 95-119

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

Wolpert DM, Doya K. & Kawato M. (2003) A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society*. B 358,593-602

Zimmerman, E., & Lahav, A. (2012). The multisensory brain and its ability to learn music. *Annals of the New York Academy of Sciences*, 1252, 179–184.