

Interaction in Musical Time

Tommi Himberg
St Edmund's College
Faculty of Music, University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy.

January 22, 2014



Declaration

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

The two experiments reported in chapter 8 were conceived, planned, conducted, analysed and reported in collaboration with Dr Marc Thompson, with equal contribution. The cooperative tapping setup used in the experiment reported in chapter 7 was piloted in student projects listed in appendix 10.

No part of this dissertation has already been, or is currently being submitted by the author for any other degree or other qualification. This dissertation does not exceed 80,000 words including tables and footnotes, but excluding tables of content, appendices, bibliography, and diagrams.

Acknowledgements

I am extremely grateful for all the help, support, and guidance I have received during this project. First and foremost I wish to thank my supervisor, professor Ian Cross for all the inspiration, guidance, and patience. I admire his ability to bring together so many different strands of research into one, coherent narrative, and to bring together such a range of wonderful people working together.

I am indebted to all my colleagues at the Centre for Music and Science in Cambridge, and at the Department of Music, University of Jyväskylä, where part of the experimental work was conducted, and the Brain Research Unit at Aalto University, where the writing up was finalised. You have shown me how research is a fundamentally collaborative and social process, and how much fun it can be when you can share the ride with awesome, smart people.

St Edmund's College became my home and people there my family—CU Finnish Society helped me at times when I felt homesick.

This thesis was financially supported by grants from Helsingin Sanomat Centennial Foundation and St Edmund's College.

Most of all, I wish to thank my family, my parents Titta and Juhani Himberg, my brothers Kai, Lasse and Joni, and my wife Janina, who have supported me through everything in life. This is dedicated to you.

Abstract

Social cognition in general, and rhythmic entrainment in particular, have previously mainly been studied in settings where isolated individuals perform controlled tasks.

Recently, a number of alternative approaches have been developed to re-define what constitutes the “cognitive system”. In addition to the individual mind/brain, the body, the social context, and musical instruments should be included in the analytical framework. In general, this means studying cognition and behaviour at settings that are as naturalistic as possible. Music and dance are ideal domains for studying these phenomena, as naturally social, embodied activities.

Extending the traditional setting poses many challenges. I make the case for focusing the analysis on the interaction of multiple participants, instead of trying to measure the performance or mind-states of the individuals. This *interactionist* approach requires a specific set of analysis tools. For this purpose, I distinguish *rhythmic synchronisation* from *entrainment* between mutually cooperative individuals. I discuss a range of options from circular statistics to cross-recurrence analysis to various correlation-based analyses.

Through a number of pilot studies, I developed a cooperative tapping setup for studying rhythmic entrainment in dyads. Using this setup, I compared human–human interaction to synchronising with a computer. Surprisingly, two human tappers reached better synchronicity than a human with a computer tapper, even though the human pairs drifted in tempo. This demonstrated the power of *mutual adaptation*.

In a second series of experiments, motion capture was used to investigate the embodied nature of rhythmic entrainment. These cross-cultural studies on African dance, illustrated in more detail how synchronicity was achieved through a process of continuous, mutual adaptation. We observed interesting contrasts in how Finnish novices and South-African or Kenyan experts exhibited embodied metrical structures.

As a conclusion, mutual adaptation is a powerful and ubiquitous phenomenon that can only be observed in real-time interactions. It is a good example of the kind of “new psychology” that can be uncovered by adopting a social, embodied, and dynamic approach to cognition.

Contents

I	Introduction and theoretical background	12
1	Introduction	13
1.1	Setting the stage	13
1.2	Approach	16
1.3	Contents of the thesis	18
2	Theoretical background	21
2.1	Key concepts and definitions	21
2.1.1	From similarity to synchrony	21
2.1.2	Entrainment vs. synchronisation	22
2.1.3	Period and phase	25
2.1.4	Rhythm and metre	27
2.2	Criteria for entrainment	29
2.2.1	Two or more oscillatory processes	29
2.2.2	Autonomy	30
2.2.3	Interaction and coupling	30
2.2.4	Entrainability or flexibility	31
2.2.5	Complex entrainment	33
2.2.6	Interpersonal interaction	34
2.3	Traditional cognitive approach	36
2.4	Embodied cognition	37
2.4.1	Embodied approach to cognition	37
2.4.2	One, three or six types of embodied cognition?	39
2.4.3	Body movement	41
2.4.4	Situated cognition: mind and the environment	41
2.4.5	Timescales of cognition: the dynamic view	43
2.4.6	Representations	45

2.4.7	Embodied vs. traditional cognition	47
2.4.8	Embodied cognition in music	48
2.4.9	Radically embodied social psychology	51
2.4.10	Summary	53
2.5	Social cognition	53
2.5.1	Historical approaches to sociality	54
2.5.2	Modern social cognitive approaches	55
2.5.3	Mimicry	55
2.5.4	Interactional synchrony and affiliation	57
2.5.5	Empathy and mirroring	58
2.6	Interactionist approach	60
2.7	Examples of interactionist experiments	63
2.8	Summary: music as an ideal domain for studying embodied, situated, social cognition	64
II	Interaction - prior research and methods	68
3	Detecting synchrony and entrainment	69
3.1	Introduction	69
3.2	Unintentional coordination	70
3.2.1	Pendulums, rocking chairs, and treadmills	71
3.2.2	Body sway and posture	75
3.2.3	Summary	77
3.3	Comparative studies on synchronisation and entrainment	79
3.3.1	Primate studies	80
3.3.2	Snowball	81
3.3.3	Vocal mimickry	83
3.3.4	Summary: comparative studies	85
3.4	Developmental studies	86
3.5	Beat deafness?	90
3.6	Summary: detecting synchrony and entrainment	92
4	Dyadic entrainment, ensemble timing, and leadership roles	95
4.1	Introduction	95
4.2	Entrainment in dyadic tapping studies	96
4.3	Ensemble and group timing	98

4.4	Conductor studies	104
4.5	Leading and following	105
4.6	Summary: entrainment and leadership in dyads and groups	110
5	Phase relationships: circular statistics and cross-recurrence	113
5.1	Circular statistics	114
5.1.1	Circular data	115
5.1.2	Mean direction and concentration measure \bar{R}	117
5.1.3	Rose histogram	119
5.1.4	Polar scatter plots	120
5.1.5	Phase conversion	121
5.1.6	Analytical methods	127
5.1.7	Rayleigh test	129
5.2	Cross-recurrence analysis	130
5.2.1	Phase space and cross-recurrence	130
5.2.2	Cross-recurrence quantification (CRQ)	132
5.2.3	CRQ results and interpretation	135
5.3	Summary: relative phase	138
6	Analysing rhythmic interaction and influence	140
6.1	Introduction	140
6.2	Pairwise intra-class correlation	141
6.3	Smoothing	143
6.4	Periodicity analysis	147
6.5	Cross-correlation	148
6.6	Granger-causality	151
6.7	Examples - mutual influence	152
6.8	Group entrainment	154
6.8.1	Phase conversion: Hilbert transform	154
6.8.2	Kuramoto model	155
6.9	Summary - analysing interaction	156
III	Empirical studies and discussion	158
7	Human vs. computer tapping partners	159
7.1	Introduction	159

7.2	Methods	160
7.2.1	Participants	160
7.2.2	Tasks and stimuli	161
7.2.3	Apparatus	163
7.2.4	Procedure	164
7.2.5	Analysis methods	166
7.3	Hypotheses	167
7.4	Results	168
7.4.1	Synchronisation	168
7.4.2	Tempo drift	171
7.5	Discussion	174
8	African dance	178
8.1	From social to social embodied cognition	178
8.2	Choir study	179
8.2.1	Choir study: background	179
8.2.2	Choir Study: participants and materials	180
8.2.3	Choir study: apparatus and procedure	181
8.2.4	Choir study: data analysis	182
8.2.5	Choir study: results	183
8.2.6	Choir study: examples of windowed analysis	185
8.2.7	Choir study: discussion	192
8.3	Dance study	194
8.3.1	Dance study: background	194
8.3.2	Dance study: methods	195
8.3.3	Dance study: results	197
8.3.4	Dance study: discussion	201
8.4	General discussion	202
9	Conclusions	206
9.1	The aims of the thesis	206
9.2	Embodiment	208
9.3	Social cognition	209
9.4	Shared manifold - a model for social cognition	211
9.5	Future directions	215

A	Pilot experiments and student projects	248
A.1	Introduction	248
A.2	Auditory and Visual Modalities	248
A.3	Delayed auditory feedback	251
A.4	Metrical Schemata	252
A.5	Effects of personality: assertiveness	254
A.6	Synchronisation and syncopation - effector comparison	255
A.7	Autism and entrainment	256
A.8	Summary	257
B	Data collection	258
B.1	Procedures	258
B.2	Using MIDI	260
B.3	Movement data	262
B.4	Audio and video	264
C	Circular statistics	265
C.1	Illustration of the cross-over problem due to tempo drift	265
C.2	Additional phase conversion methods	266
C.2.1	Group sync: Abstracted beat as the referent	266
C.2.2	Polyrhythmic relationships	266
C.3	Additional details on circular measures	267
C.3.1	Uniform probability plots	267
C.3.2	Direction and spread of sample	267
D	Cross-recurrence parameters	270
D.1	Parameter selection	270
D.1.1	Delay	271
D.1.2	Embedding dimension	271
D.1.3	Threshold	273
D.2	Example: parameter selection in improvisation trial	273
E	Publications	275
E.1	Refereed articles	275
E.2	Conference proceedings papers	275
E.3	Conference proceedings (as editor)	276
E.4	Invited and keynote lectures	277

E.5	Conference presentations and posters	277
E.6	Teaching materials	279

List of Figures

2.1	Mass on spring	26
2.2	Weakly coupled metronomes	32
5.1	Onset on unit circle	118
5.2	Theta and R	118
5.3	Rose histogram	120
5.4	Polar scatterplot	121
5.5	Rose histogram comparison	123
5.6	Phase conversion methods	126
5.7	Phase space	131
5.8	Recurrence plots	133
5.9	Markers in communication experiment	135
5.10	Cross-recurrence analysis example	137
6.1	Intragroup correlation visualisation	144
6.2	Examples of smoothing	145
6.3	Windowed cross-correlation, example	149
6.4	ITI and cross-correlation between experimenter and participant with ASD.	153
7.1	Structure of tapping trials	165
7.2	Polar scatterplot of synchronisation	169
7.3	Synchronisation - effect of <i>Partner</i>	170
7.4	Synchronisation - effect of <i>Task</i>	171
7.5	Polar scatterplot of tempo drift	173
7.6	Tempo stability - effect of <i>Partner</i>	174
7.7	Tempo stability - effect of <i>Task</i>	175
8.1	Pair-wise correlations	183

8.2	Cross-correlations, vertical	184
8.3	Cross-correlations, horizontal	186
8.4	Cross-correlogram, African song	187
8.5	Cross-correlogram, Finnish song	188
8.6	Order parameter	189
8.7	Order parameter, African song	189
8.8	Phase deviation	190
8.9	Vertical head and foot acceleration	191
8.10	Hypotheses	195
8.11	Motion capture image	197
8.12	Windowed cross-correlations of head centroid velocities in ver- tical dimension.	198
8.13	Rose plots of head markers	199
9.1	Shared manifold	212
A.1	Synchronisation and coordination errors	250
A.2	Average inter-tap intervals of the beats in the bar	253
C.1	Linear to Circular	265
C.2	Uniform probability plots	268

List of Tables

6.1	Intraclass correlation	143
7.1	Summary of tapping partners	162
7.2	Factorial design	167
7.3	Summary of effects in ANOVA	176
8.1	\bar{R} -values of the pair performances - head and root centroids and right hand marker.	201
8.2	Effect sizes of the contrast analysis.	201
A.1	Correlations of relative velocities and metrical hierarchies . .	254

Part I

Introduction and theoretical background

Chapter 1

Introduction

“I’m picking up good vibrations
She’s giving me excitations”

- Mike Love/Beach boys: *Good Vibrations* (1966)

1.1 Setting the stage

After learning to sit straight and locating the middle C, a young, aspiring pianist will need to learn two timing skills that are crucial for any music performance. One is the ability to internalise a regular beat, a pulse that forms the spine of a musical piece, the *when* to the flesh of *what* is being played. The second skill is to entrain his or her motor actions to that beat, to press keys in time with the internal pulse. Very quickly another skill is required, as coordinating the actions of the left and the right hand is learned. Very soon, interpersonal entrainment is introduced, as our little piano player gets to play with others. From the first time ‘Twinkle, twinkle little star’ is played to proud grandparents, the young performer is engaging in musical communication and experiencing emotions and eliciting them in the audience. From the beginning, the music performance shapes to be a profoundly social, emotional and intellectual experience, bringing people together, entraining them into an interaction in musical time.

Most likely, our little pianist will learn to play Western classical music, which is a prime example of a *presentational* musical practice (Turino, 2008). Social aspects are important for understanding the psychological processes related to presentational music, but they are crucial for understanding

what Turino calls *participatory* musics. These are musical cultures and genres where the distinction between performers and listeners is blurred; where even the *forms* of music invite joining in, participating in music-making—communicating. In these practices, the function of music-making is to enhance social bonding, and music itself grows from the interactions, rather than existing as abstract, pre-determined pieces of music (Turino, 2008, 36). These musical practices are often aurally transmitted, and children “grow” into them rather than receive formal training in them. Our current research on music cognition is strongly based on our own Western “high” traditions of classical and jazz music, which leads to a bias where we see music as much more fixed, symbolic, and cerebral than what it actually is.

There is a long tradition in Western musicology to study music in this disembodied, symbolic form, and from that perspective it is understandable that a similar approach would dominate also the study of music psychology. And given the peculiar characteristics of Western classical music (where—stereotypically—specialised performers reproduce complex, pre-written pieces of music to passive listeners), the limitations of this approach are not so obvious as they are when looking at any other music culture around the world, or indeed those features that different music cultures share. For a theory or model of music cognition to have general relevance, it should strive for universality, or at least wide applicability across different musical genres, traditions and cultures. Therefore, it is necessary to employ a definition of music that does not reduce it to mere auditory stimuli that are passively perceived (I. Cross, 2009, 2012).

Music intentionalises time. It generates expectations of both *what* is to happen, and especially strong ones regarding *when* that is going to happen. So much so that while we might not be able to sing the melody line or play the piano part, we can effortlessly tap in time with the music, predicting perfectly when the next note is going to take place. In fact, it is even difficult not to. Our body starts to move automatically when, say, ‘Good Vibrations’ plays. This is a general musical capacity we all have (apart from very few exceptions, see chapter 3.5). Music is a human universal, in that every culture has what can be identified as “music” (see Blacking, 1973), and almost everyone in those cultures is able to engage with music, even though they might not have special musical training. The ability to engage with music develops very early in childhood, with the sensitivity to musical elements starting

to develop already before birth, alongside emotion recognition and basic perceptual abilities needed for language comprehension (Huotilainen, 2010; Trehub, 2003; Winkler, Háden, Ladinig, Sziller, & Honing, 2009). This can be seen in the ubiquitousness of lullabies across the world, and how lullabies “work” even for pre-term babies (see e.g. Farhat, Amiri, Karbandi, Esmaily, & Mohammadzadeh, 2010; Standley, 2002).

Music is social. Even though we can enjoy music by listening to it in the comfort of our own homes, or use it to drown out sounds of the environment (like I am doing right now, walling myself from the rest of the people in this cafe by listening to Bach’s cantatas through headphones), music is made in social interaction, for social communication¹. Becoming a musician can be a solitary pursuit with endless hours spent practicing alone, but at its core, beyond these cultural idiosyncrasies, music is a social activity, which is about communicating emotions and ideas to others, and about engaging in entrainment and joint action with others. Music is often performed in ensembles, and part of becoming a good ensemble musician is to become an expert listener, picking up the subtlest of cues and learning to anticipate the actions of others, and learning to influence others with one’s own actions, and in turn be influenced by the actions of others. The joint action of performing a piece of music, or improvising music together requires a large constellation of skills and abilities (Schellenberg, 2004).

Music is embodied. For there to be musical sounds, there needs to be movement. The sounds of music can be generated in many ways, by hitting the skins of drums, by creating controlled vibrations of vocal chords, reeds or lips with aspirations of air from the lungs or by plucking or bowing strings, or pressing piano keys to launch hammers at strings. Through musical training, musical instruments become parts of the musicians’ bodies, and extensions of their somatosensory and motor systems. Musical thinking is not abstract manipulation of symbols, but a combination of imagination and physical routine; mental images and motor programmes.

Music is pleasurable. Even sad music gives us pleasure (Vuoskoski,

¹In this era of recorded music, music of course exists also in a “dematerialised” form where no social interaction is necessary between performers and listeners (see Auslander, 2008). Although recording technology poses interesting questions about “how music works”, this is a recent development (especially in evolutionary terms), and therefore characterising music as social interaction provides a more universal viewpoint into the musical mind.

Thompson, McIlwain, & Eerola, 2012), and music is a great tool for mood regulation, usually towards more positive moods (Saarikallio & Erkkilä, 2007). Music brings people together and generates strong experiences of belonging. Music is a great tool for therapy, especially for issues related to social interaction (Bruscia, 1987; Gold, Wigram, & Elefant, 2006; Pavlicevic, 1997; Wigram & Gold, 2006). Music is communication, but it is not a language: it is “safer” and more inclusive. You can take an active part in music-making without having to be the centre of attention, while in a conversation, taking an active part usually requires one to speak alone in a group, being the focus of everyone else’s attention. Also, in music you do not need to commit to declarations of fixed, propositional meaning, but you can engage with others under a cloak of ambiguity and flexible intentionality. While music can touch our innermost feelings and convey our most intimate ideas, this fluidity shields us from them being exploited. The feeling of connectedness, however, is an enriching consequence of musical interaction and the inherent reward that music brings, transforming the abstract into the concrete, visceral domain.

Music seems to be almost magical, and in the writings of musicians, composers or music scholars, often something ethereal and unfathomable. But, it is also concrete, physical, and real, and within the realm of scientific study. Musical behaviours and abilities can be studied just like any others, using tools from psychology, philosophy, musicology, cognitive science, computer science, neuroscience et cetera (given the complexity of the phenomenon, a multi-disciplinary combination of these would be preferable). The social, embodied, emotional, communicational nature of music can be taken into account and even made the focus of the study, as I hope to demonstrate.

1.2 Approach

In this thesis, I am taking an embodied, social cognition approach to studying music performance. This entails paying attention to both the biological roots of musical abilities (our capacity for music), as well as the cultural practices of music. These two sides are not separate but entangled in a delicious fashion in music. They are combined in for example the study of evolution and music (I. Cross, 2001, 2008; Wallin, Merker, & Brown, 2000). Studies in evolution and music can address questions such as which

cognitive functions and skills comprise the music faculty, how this faculty is related to others (such as language) and what the music faculty is *for*. The evolutionary approach is also healthy in reminding one about the importance of having a wide and clear enough definition of music itself. Music is not just notes on paper or bits on a CD, but social, embodied action. Also, it is important to remember that the biological and cultural approaches do in reality go hand-in-hand, and whichever your point of view, the other needs to be recognised (see e.g. Cacioppo, Berntson, Sheridan, & McClintock, 2000).

Seeing music as a social, embodied activity means that music is something that we *do together*, by moving together or in order to move together. “Musicking” is an excellent term for this (Small, 1998). Evolutionary theories about the origins of music often stress the social aspects of it: musicking provides a way to promote group coherence, serve as a way of bonding between a mother and a child, or as a form of communication (Huron, 2001; Perlovsky, 2010). One of the key characteristics of music is that it unfolds in time, it structures and intentionalises time. This temporal nature of music will be the focus of this thesis, which means that it is important to study the dynamics, the temporal development of musical activities.

A lesson learned from dynamic systems studies is that the whole is often more than just the sum of its parts: new phenomena *emerge* from interactions of the individual components of the system. The focus is shifted from the internal processes of the individual to the relationship of the pair—how they act together, how their relationship evolves over time. As this relationship is absent when studying isolated individuals, it can not be extrapolated from just studying a number of individuals and adding that data together. The smallest possible unit of analysis for social cognition is a dyad (Semin & Cacioppo, 2008). The shift of focus from individual to social aspects of music-making also requires a fundamental shift in methodology, especially in measurements and analysis methods. Instead of measuring static, individual states, we need dynamic measurements of interactions and influence. Working on suitable methods has been a major part of this project, and discussing them forms a bulk of this thesis.

This thesis is primarily a study on entrainment—the temporal, mutual synchronisation of two or more participants (M. Clayton, Sager, & Will, 2004). Studying entrainment thus requires extending the view from the individual to the dyad or group. Entrainment is also embodied, as mu-

sis performance (usually) requires the coupling of auditory inputs to motor outputs. Entrainment is not restricted to music, rather it is a very basic physical phenomenon that can occur in many contexts and systems, from neural or atomic levels to groups of people, to even celestial bodies (Strogatz, 2003; Toussaint, 2013). It has also many applications in human sociality, as it underlies all interpersonal communication regardless of its modality. It is largely subconscious but can also be consciously employed and modulated (Thaut, 2008).

1.3 Contents of the thesis

In the next chapter, I will introduce and define the key terms and present a theoretical framework for the study. This framework consists of embodied and social cognition theories and a recent body of research in music performance, entrainment and musical communication. The social and embodied approaches to cognition have gained momentum recently, even though the idea that our bodies and social environments affect our thoughts and behaviour are ancient. The core ideas of the embodied approach to cognition were discussed by William James (James, 1890); his *ideomotor principle* forms the basis of current theories of action-perception cycles and their shared representation and the recent theories about the mirror neuron system.

In chapters 5 and 6, I will discuss previous studies on dyadic and group entrainment, as well as demonstrate a number of measurement techniques and research methods for studying interpersonal interaction in a musical context. In terms of statistical methods, I will introduce two families of methods in more detail: *circular statistics* (Fisher, 1993) and *cross-recurrence analysis* (Marwan, Romano, Thiel, & Kurths, 2007). In addition, I will present methods for studying interaction and mutual influence in dyads and groups. I will discuss both discrete data (such as tapping data) and continuous data (such as movement data).

I will present empirical studies in two chapters and one appendix. In chapter 7, I will demonstrate that interpersonal entrainment differs from individual synchronisation by reporting results from a dyadic tapping study. This study compares tapping with another human to tapping with non-responsive computer tappers. The data was collected in two stages; the first

data collection took place in Cambridge in 2006 and the second in Jyväskylä in 2009, with an enhanced setup. These results have been presented in conferences (e.g. ICMPC 9 conference in Bologna, Italy (Himberg, 2006) and in the 16th Annual Symposium of Music Scholars in Finland (Himberg, 2012)).

In chapter 8, I will report two motion capture studies, thus extending the embodied view from finger-tapping to full-body movement. These studies were conducted in Jyväskylä in collaboration with Dr. Marc R. Thompson. These studies are cross-cultural dance studies exploring the role of expertise in embodying metre, and of dyadic and group entrainment. In the first part, we studied group entrainment and movement synchrony in a mixed South-African–Finnish choir, comparing experts and novices. The second part is a study on dyadic entrainment in Kenyan dance, featuring participants from Kenya and Finland.

The studies in chapter 8 have been published. The choir study has been published in *Dance Research*, co-authored by Marc Thompson and myself (2011, vol. 29, number 2, pp.305–328)². We shared first authorship as we contributed to the study and the paper equally. The dyadic dance study was a pilot study that has been presented as a poster in ICMPC 11 conference in Seattle in 2010, and published as a proceedings article, co-authored by Marc Thompson (Himberg & Thompson, 2010).

In the concluding chapter, I will summarise the findings and outline some future prospects in the field of embodied, social cognition, and especially how studying music and dance can be helpful in discovering how the embodied, social mind works.

Preceding these studies, I conducted a number of smaller pilot studies, in order to develop the cooperative tapping setup used in the human vs. computer partner experiment. These were mainly exercises in experimental design and data analysis. The research questions posed in these studies are still valid, though the data so far is not solid enough to warrant drawing firm conclusions. These pilots are briefly presented in appendix A.

This project started in 2003, when most of the literature on dyadic entrainment, or even embodied and social cognition had not been published yet. The only example of actual, interpersonal entrainment study that I

²N.B., the article formed a part of Marc’s PhD Thesis that was accepted in the University of Jyväskylä in March 2012.

could find was the study by Mates, Radil, and Pöppel (1992), which became the basis of my own experimental design. The development of the measurements, writing software to analyse dyadic tapping data, designing, running, and analysing the experiments, and finally writing them up has taken a long time. While my work has been in progress, many researchers have identified the same problem and started working on dyads and groups, and now there is a core literature that is converging in terms of findings and methods.

Chapter 2

Theoretical background

2.1 Key concepts and definitions

2.1.1 From similarity to synchrony

Behavioural coordination takes place in many domains, including verbal and non-verbal communication (e.g. speech rate convergence and gesture similarity) to motor behaviour (e.g. mutual adjustment when carrying planks together). Thus there are many kinds of manifestations of behavioural coordination. Bernieri & Rosenthal (1991) make a distinction between two main types: behavioural matching (such as mirroring the conversation partner's gesture, converging in vocabulary or speech rate) and interactional synchrony (such as gait entrainment or making music together). The main difference between these two groups is that behavioural matching refers to the (relative) *simultaneity* of events, such as certain gestures in conversation (rubbing chin, crossing legs), or *convergence* or *rapport* of features such as speech rate, syntactic features, posture or facial expression (Condon & Ogston, 1967; Gonzales, Hancock, & Pennebaker, 2010; Kendon, 1970; LaFrance, 1979). In contrast, *interactional synchrony* can be characterised as convergence between sustained rhythmic processes. This temporal coordination of continuous processes has been called *synchronisation* as well as *entrainment*; I will discuss possible distinctions between the two terms in more detail in the next section. However, while sometimes the one-off matching has also been called behavioural *synchrony*, it would not be possible to refer to anything in Bernieri & Rosenthal's first category as *entrainment*.

This field is a cornucopia of terms. Giles, Coupland & Coupland (1991)

reviewed studies on “convergence” and “divergence” in conversations, listing for example the many linguistic features that converge between participants in conversations, from speech rate to choice of syntactic structures to posture. In their review, they noted that many terms have been used in the literature to describe such convergence, including *accommodation* (their own choice, cf. communication accommodation theory, CAT), *reciprocity*, as well as *synchrony*. In describing these one-off events and the increase of their *similarity*, I would avoid the term *synchrony* to avoid confusion. There are still plenty of terms to choose from, for example *rapport*, *matching*, *mirroring*, *alignment*, or as above, *convergence* or *correspondence*. Thus both *synchronisation* and *entrainment* would be reserved for continuous, (quasi-)periodic processes. These two have subtle differences in meaning, which I will discuss in the next section.

2.1.2 Entrainment vs. synchronisation

In the literature, there is some confusion regarding what *synchronisation* and *entrainment* mean, and whether they are synonymous or not. Rhythm and timing studies have almost exclusively looked at individuals synchronising with a metronome. I will demonstrate in this thesis that two people mutually synchronising with each other is phenomenally very different from the one-sided case, and thus it would make sense if a separate term, *entrainment*, was reserved for it.

“Entrainment describes a process whereby two rhythmic processes interact with each other in such a way that they adjust towards and eventually ‘lock in’ to a common phase and/or periodicity.” (M. Clayton et al., 2004, 2). Along similar lines, Pikovsky, Rosenblum, and Kurths (2001) state that “This adjustment of rhythms due to an interaction is the essence of synchronization.” (p. xviii). Both emphasise that there are further criteria that must be met for a phenomenon to be classified as entrainment. According to these criteria, there needs to be two or more oscillators, or rhythmic processes, that are independent, interacting, and flexible enough to converge in their phase or period towards the other(s). I will return to these criteria in more detail in the following section.

Pikovsky et al. work on synchronisation in general, and use the two terms interchangeably to refer to co-adapting, interacting, continuous processes within and between various physical systems (electric circuits, liquids,

celestial bodies, fireflies ...).

Looking at synchronisation/entrainment of conscious, intentional agents, there are a few possible distinctions. One is that *entrainment* refers to the general phenomenon that is or is not expressed in overt *synchronisation* of movements. Here, a person could be entrained to the beat of music even when they are sitting completely still, assuming that they hear the music and in their brains some circuits are firing in a pattern that is related to the pattern of beats in the music, allowing the person to engage in synchronous movements if they wish to do so. Here, synchronisation is the specific behavioural outcome of being internally entrained to the external stimulus.

Another option has to do with the number of agents capable of adapting their behaviour: one adaptive agent would be *synchronising* with an external, non-responsive referent (such as a metronome or the circadian rhythm), while two or more adaptive agents would be *entraining* to each other. In the latter case, all rhythmic processes that are included, would, as Clayton et al. say, “interact with each other in such a way that they adjust towards and eventually ‘lock in’ to a common phase and/or periodicity.” The key difference is that in the case of one-sided synchronisation, only one of those processes is capable of adjustment, and therefore calling that *entrainment* would perhaps be a stretch.

To account for the different number of agents and processes, there are some special terms used already. For instance, synchronisation or entrainment *within* the individual is sometimes called *self-synchronisation* or *self-entrainment*, for example when the movements of an individual’s left and right arm are synchronised. This was the focus of the classic study on inter-limb coordination by Scott Kelso (1984). Similarly, sometimes the case of multiple people entraining to each other is called *mutual synchronisation*, as for example by Pikovsky et al., although Clayton et al. refer to this just as *entrainment*, as mentioned above. Yet another complication arises from different kinds of synchronisation, as sometimes a distinction between period and phase synchronisation is made. As it is a feature of many systems that once they get synchronised, they stay that way, entrainment could also be called period or phase *locking*. And indeed, Pikovsky et al. clarify (p. 22–23) that they use the terms *locking* and *entrainment* interchangeably, as synonyms to describe what happens in the synchronisation process.

Most of the previous research in the field of rhythm perception and pro-

duction has been based on experimental designs where individuals are synchronising with non-responsive referents, and the term *synchronisation* has been used consistently for this process. In my opinion, it is therefore logical to adopt a different term to be used in the emerging field of research, which studies the process of *entrainment* in dyads or in groups¹. One of the aims of this thesis is to demonstrate that synchronisation and entrainment, as defined here, are psychologically different phenomena. Thus, from the point of view of cognitive research, this distinction is a valid one, even if for someone studying synchronisation in purely physical contexts (mechanical, chemical, etc.), this distinction might be of less value.

There are also other nomenclatures. Phillips-Silver and her collaborators recently suggested that a case where one “entrainable system”, or a social agent, gets someone else’s rhythmic output as their rhythmic input, and then entrains, should be called “social entrainment” (Phillips-Silver, Aktipis, & Bryant, 2010). So, for instance, a case where one person taps to the music played by another person (but where this tapping has no effect on the person playing the music), would be social entrainment. This is very close to the one-sided synchronisation case in dyadic tapping studies (see e.g. Konvalinka, Vuust, Roepstorff, and Frith (2010) and chapter 7 in this volume). For Phillips-Silver et al., the case where both “entrainable systems”, or social agents entrain to each other (the rhythmic output of one is the rhythmic input of the other), is called “mutual social entrainment”. And finally, when there are more than two, that is called “collective social entrainment”. (Phillips-Silver et al., 2010). I find the first category problematic, as in my view it is not social or entrainment at all. While Phillips-Silver et al. say that it is the same as “interpersonal entrainment” elsewhere (referring to Clayton et al.), it really is only one-sided, one person synchronising to an external referent that does not connect back to them in any way. And in social cognition, this feedback connection, mutuality and symmetry of the relationship is crucial.

For example, *joint attention* is a very basic mechanism of social cognition (Knoblich & Sebanz, 2008). It differs from the mechanistic *gaze following* (a much more common and even more “basic” social skill), in that it is

¹In mechanics, entrainment is used also in situations where one oscillator is driven by another, fixed one. In this situation, only one of the oscillators can change, so in the parlance of this thesis, this should be called synchronisation instead.

“joint”: there is a mutual recognition that both are attending to the same object. This is crucial, because while gaze following is a survival strategy that helps agents to be aware of potential threats or treats in their environment, joint attention is a *social tool* for influencing other social agents and communicating intentions, wants and wishes (Carpenter, Nagell, & Tomasello, 1998). Similarly, synchronisation is a mechanical way of connecting with periodic changes in the environment, akin to circadian synchronisation to daylight cycles, while entrainment, with its mutuality, is a process of social cognition and a tool that underlies interpersonal communication, intentionality and emotional processes. In short, I argue that *joint attention is to gaze following as entrainment is to synchronisation*.

To summarise, a person, trying to tap with a metronome, is an example of the process of *synchronisation*. Two people tapping with each other is an example of *entrainment*. This distinction might seem small, but in a “gibsonian” parlance (J. J. Gibson, 1977), the *affordances* of these two situations to the participants are very different, as entrainment affords interaction while synchronisation does not. And although entrainment can occur between physical objects, the focus and interest in this thesis is on the psychological and cognitive aspects of these processes, and what they facilitate in terms of interpersonal communication and interaction.

2.1.3 Period and phase

For all the consequent analyses and methods it is important to understand the basic concepts of period and phase. To illustrate these concepts, let us look at *simple harmonic motion*, of which one example is a mass on a string (see figure 2.1). When this idealised system (no damping due to friction or drag) is in equilibrium, the mass hangs from the string at the height of the dashed equilibrium line, not moving. But when it is displaced—pulled down—and then let go, it starts bouncing up and down in simple harmonic motion. The force of the string pulls the mass up towards the equilibrium, and overshoots it due to the momentum the mass has gained. Above the equilibrium line, the downward force of gravity exceeds the upward force of the string, so the mass starts to decelerate, then it momentarily stops and then starts to fall. Again, momentum makes it overshoot, and the oscillation continues.

In this system, the period is the time it takes the weight to complete one

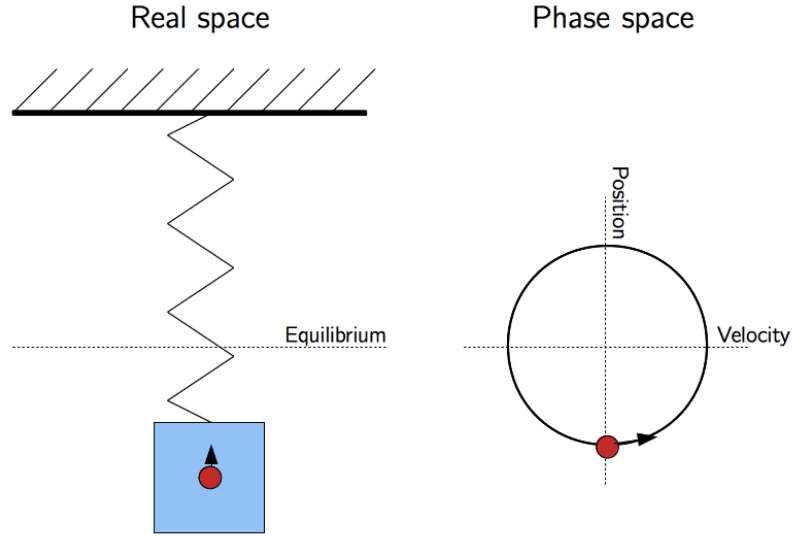


Figure 2.1: Mass on a spring, a view in the real space, when the mass is at the lowest point of its oscillation, and a phasor view which shows the periodicity of the movement.

cycle from, for example, the bottom position back to the same place. The phase space of this system is depicted in the right panel of figure 2.1. This phasor view shows that the system can be depicted as a uniform circular motion. The phasor completes a full circle during each period. The phase of the system at any moment i can thus be defined by locating the point (x, y) (where x = velocity, y = position) along that circle, and then quantified as the angle between (x_i, y_i) and (x_0, y_0) . Looking at the phasor we can see that this angle θ grows linearly until it reaches 360 degrees or 2π radians, then it returns to zero and starts to grow again.

To move this theoretical example from classical mechanics to the realm of music, imagine the mass on a spring plays a tone every time it reaches the bottom position, as if plucking a string of an air guitar. The resulting sound for the simple harmonic oscillator is a series of regularly timed, uniformly spaced (isochronous) notes. Here we can see that in musical terms, the period is thus related to the tempo of music, each period representing an inter-beat

interval or the time between two successive beats. In this example, as in the case of a metronome, this inter-beat interval is constant, as we can say the period of oscillation of our mass on a spring is constant. In other words, the pulse is *isochronous*.

To say two processes (for example, two virtual air-guitarist masses on ideal undamped springs) have the same period means they have the same tempo, or in other words their inter-beat intervals are the same, however they might still be out of phase. For instance, imagine the beats are crotchets, but instead of the note onsets of the two players occurring simultaneously, one plays their notes in syncopation, consistently a quaver later. Thus the second player's note onsets take place half way through the period of the first player's beats, or 180 degrees or π radians out of phase. This is called *anti-phase synchrony*. These concepts of period and phase form the basis of circular statistical analysis, and will be examined in more detail in chapter 5.

2.1.4 Rhythm and metre

“Meter is a musically particular form of *entrainment* or *attunement*, a synchronisation of some aspect of our biological activity with regularly recurring events in the environment” (London, 2012, 4).

Metre provides a structure that allows us to not only track the rhythm and timing of the music, but also predict the temporal course of events. As Justin London puts it, “meter provides a way of capturing the changing aspects of our musical environment as patterns of *temporal invariance*” (p.5, emphasis in original). Metre organises the rhythmic patterns we hear into a structure and provides a hierarchical and sequential structure for them. This “temporal invariance” means that we can perceive two phrases or rhythmic patterns as the same, even if there are small expressive timing differences between them.

While rhythm refers to the pattern of musical events, metre is an organising principle, a structure. Metre is a cognitive concept, even though it is often seen as a musicological concept, as a feature or a characteristic of a piece of music (it goes in $\frac{4}{4}$ metre, or $\frac{3}{4}$ metre etc.). Metre is hierarchical, in that it consists of different layers or levels (Large & Jones, 1999; Lerdahl

& Jackendoff, 1983). Metrical positions at one level are subdivided on the next level down, and form larger structures together at the higher level. We can entrain in a flexible way to different metrical levels and can switch our attention between different levels at will. For example, we can clap in time with music but shift from clapping on every beat to clapping on every other beat, which corresponds to shifting to synchronising with a higher metrical level.

Metre is important for perception of music, it forms a grid that allows us to predict the timing of future events (Huron, 2006; London, 2012; Rohrmeier & Koelsch, 2012). This metrical grid allows us to direct our attention to important events, and depending on our intentions and needs, either analytically to individual events or in a future-oriented way to the overarching musical structures (Jones & Boltz, 1989; Large & Jones, 1999).

At the core of joint action is the concept of intentionality. When Cross says music “intentionalises time in sound and action” (I. Cross, 2005), he means that music, and essentially *metre* in music (alongside tonality, and melodic and harmonic contours and progressions) provides a flexible but clear platform for joint attention, collaboration and other social processes. Tomasello and colleagues (Tomasello & Carpenter, 2007; Tomasello, Carpenter, Call, Behne, & Moll, 2005) write about *shared intentionality* (see section 2.5.2), as being fundamental to complex collaboration. They suggest that as a social capacity it is uniquely human, and entails not only the ability but also the motivation to share mental states. According to Tomasello et al., this shared intentionality (motivation to share intentions and goals in order to make them common ones, and the means to coordinate actions towards these goals) also entails shared representations, as the collaborators need a shared “map” of reality, a way to jointly guide progress towards the goals. In music making, metrical schemata can be seen as such shared representations². Co-performers need to entrain to each other, but in the absence of a metronome, they instead entrain to a pulse that they co-create together, and a metrical structure they agree upon.

Metre is embodied, in that it is not just an abstract psychological concept but something that we act and represent in our body movements. A simple example of this is tapping one’s foot while playing a piece of music on an instrument. “Storing” the beat and the tempo to relatively large muscles, it is

²Although later I will discuss whether mental representations exist or not.

possible to focus more attention to the fine motor acts needed to produce the musical texture. I will give further examples of embodied metre in chapter 8 using full-body movement data, showing differences between how experts and novices in African dance embody metre and rhythm. Also, I will present some data from tapping studies in chapter 6, looking at timing within musical bars, strong and weak beats etc.

2.2 Criteria for entrainment

Formulating clear criteria for entrainment is important, as the first objective in an analysis of interaction is often to determine whether there is entrainment or not. Sometimes it can be difficult to demarcate between actual entrainment and apparent simultaneity or synchronisation by chance.

Following Pikovsky et al. (2001) and Clayton et al. (2004), I propose that the criteria are: 1) there should be two or more oscillatory processes; 2) these processes are autonomous; 3) they interact with each other; 4) period and/or phase convergence should be observed. I will first discuss these general criteria for entrainment as a physical phenomenon, and then continue on to analyse what these mean for the special case of interpersonal entrainment, looking at a) agency; b) mutuality; and c) flexibility.

2.2.1 Two or more oscillatory processes

First of all, in order for there to be synchronisation or entrainment, we need temporal order. Not just simultaneity, as in a transient sync remnant of things happening at the same time because they have the same triggering cause (Steven Strogatz used the example of a car exhaust making a bang and scaring of a flock of birds (2003, 262)). Rather, we look for continuous processes that oscillate, or periodically vary in value. For example, this could be a swinging pendulum (whose displacement from the vertical zero-line changes in a sinusoidal fashion) or a person clapping hands or tapping a drum, where the sound onset generates a “spike” to otherwise steady zero-state (no sound). Obviously, for entrainment, at least two such processes are needed. To keep things simple, I will now describe the case of two processes, but the criteria can be applied to a larger number of processes, as well.

2.2.2 Autonomy

To qualify as entrainment, the two (or more) processes must be autonomous. This means that the processes (or oscillators, if you will) need to have their own energy source that allows them to oscillate even in the absence of the other, or in other words, they need to be self-sustaining (Pikovsky et al., 2001; Strogatz, 2003). This could be a metronome that gets its energy from its wound-up spring, or a person playing a rhythm on a drum or walking, as long as they are able to sustain their rhythm for a relatively long time.

From the point of view of physics, the autonomy-demand is to make a distinction between *entrainment* and *resonance*, where the resonating object gets all of its energy from the external source, and will stop as soon as the source is taken away, or after a short transient (the situation in the opening *Beach Boys* quote is likely resonance). This is the case when for instance a guitar string starts to resonate in response to the air pressure changes generated by another string with a related eigenfrequency vibrating near by. Entrainment, in contrast, is observed when two processes would exist regardless of each other, but through interaction they converge in terms of the phase and/or period of their oscillation/vibration.

It is also possible that two sets of data look synchronous, because they are produced by the same process. For instance, the displacement and velocity of a pendulum are perfectly in sync, but they can not be said to be entrained, because one is dependent on the other; both are measurements of the same process (this example was used by Pikovsky et al. (2001, 15)).

In a sense, this condition is analogous to evaluating causality between experimental variables. If two variables are both caused by the same, confounding variable, their correlation does not come as a surprise, and of course there is no causal link between the two.

2.2.3 Interaction and coupling

Not all processes that appear synchronous are entrained. They might have the same period and even phase, but if they are not connected to each other, and can not influence each other, this synchrony is just a coincidence, not due to synchronisation or entrainment ³.

³Terminologically, we can talk about the processes being *synchronous*, without (the process of) *synchronisation* or *entrainment* taking place.

This brings us to the third condition, according to which, for there to be entrainment, the oscillators need to be coupled. In other words, they need to have a connection and be able to mutually influence each other. For example, two musicians in separate, sound-proofed rooms without means of communicating with each other can not entrain, and any synchrony between them is merely coincidental. But if they can see, or even better, hear each other, they can entrain to each other, play in time, reacting to each others' timing cues.

But this coupling or connection should not be too weak or too strong. Let us take the example of two metronomes. If the metronomes are placed on a surface that can move laterally (to the direction of the swinging of the pendulums of the metronomes), for example on a plate that is placed on rolls (see figure 2.2), the movements of one metronome can be communicated to the other via the plate movements, and the metronomes can entrain. The moving plate couples the two oscillations to each other, allowing the metronomes to mutually influence each other's movements. However, if the metronomes are placed on a rigid table that can not transmit the movements, the coupling is too weak for entrainment to take place. Coupling can also be too strong. If the pendulums are connected with a rigid beam, this forces them to move in synchrony, but this would violate the condition of autonomy: the two pendulums have no other option but to move together, making the two metronomes essentially just one system.

2.2.4 Entrainability or flexibility

To ensure that the synchrony of two processes is not merely a coincidence, their ability to entrain needs to be proven. The processes should be able to find a common period and lock in phase even when they start a small distance (in frequency and/or phase) apart — they should have an *entrainment region*, within which they will eventually find a common pulse. If the frequencies of the two processes are f_1 and f_2 , and their difference is Δf , and when coupled, their frequencies respectively F_1 and F_2 , and the frequency difference in that coupled system is ΔF , then if there is entrainment, the ΔF should over time become zero for some (small) values of Δf . This means that even when the two originally have different periods, they adjust to a common one when coupled.

This distinguishes true entrainment from the case where we would ob-

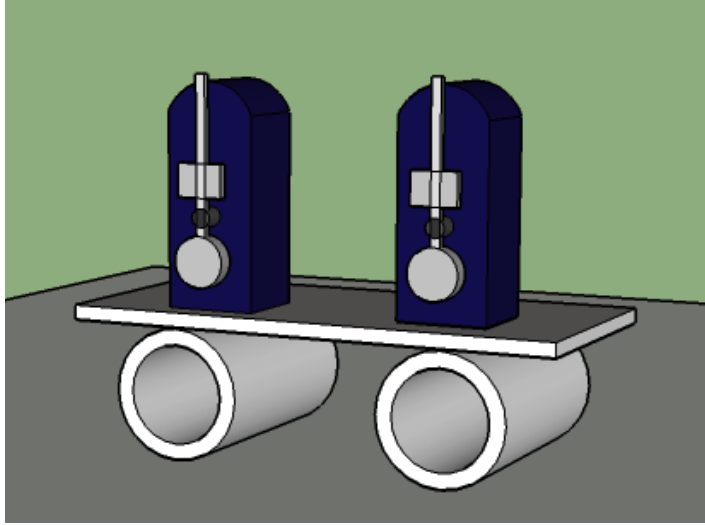


Figure 2.2: Two metronomes, weakly coupled through the plate that can move laterally, transmitting the movements of the pendulums to each other.

serve “phase locking” of two processes that would have this constant phase relationship simply as a virtue of having exactly the same period. If we observe two processes that are synchronous, we could test for entrainment by perturbing one of the two. In the case of coincidental synchrony, this introduced phase or period error remains, but if the two are entrained, the error will be corrected soon (provided that the perturbation is not too large to put the Δf outside the entrainment region).

A recent example of using this tactic to test for synchronisation comes from Aniruddh Patel and his collaborators (Patel, Iversen, Bregman, & Schulz, 2009) who wanted to check if the sulphur-crested cockatoo ‘Snowball’ is actually synchronising with the piece of music that it is moving with in the famous YouTube-video⁴, or whether its preferred frequency of movement simply coincides with the tempo of the music. In order to test this, the researchers constructed different versions of the piece, ones that were slower and faster than the original, but otherwise identical. In their studies they observed that the bird did to some degree speed up its movements in response to the faster version, but was unable to follow the slower versions.

⁴The original video has got over 5 million views at <http://www.youtube.com/watch?v=N7IZmRnAo6s>.

I will return to this example and review the methodology and analysis of Patel et al. in the next chapter.

2.2.5 Complex entrainment

For entrainment to take place, the two processes do not need to be identical, and they don't need to have identical periods or even a particular phase relationship. Above, the periods of the two entrained processes have been in a 1 : 1 relationship to each other, but more generally, they can be in a $m : n$ relationship to each other. In music, relationships where either m or n is 1 and the other a small integer are very common, and reflect entrainment on another level of metrical hierarchy. Relationships where both m and $n \neq 1$ are often called polyrhythms. As m and n get larger, the stability of the synchronisation very quickly diminishes. Even simple polyrhythms are less stable than the prototypical 1 : 1 or simple 2 : 1 or 3 : 1 synchronies (Deutsch, 1983; Handel & Lawson, 1983; Povel, 1981; Summers, Todd, & Kim, 1993; Vuust, Roepstorff, Wallentin, Mouridsen, & Østergaard, 2006).

This definition of entrainment does not prescribe a particular phase relationship for the two (or more) processes. This means that the phase difference of the two processes, $\Delta\phi$ can have other values than zero, though $\Delta\phi = 0$, or *in-phase synchrony*, is perhaps the most stereotypical case. For example, when two people walk in-phase, their feet strike the ground at the same time – left with left, right with right. In *anti-phase synchrony* the steps still occur at the same time, but this time the left foot step of one walker coincides with the right foot step of the other. As phase is defined to be 2π -periodic, or to grow by 2π radians or 360 degrees during each period of the oscillation, the phase difference ($\Delta\phi$) of the anti-phase synchrony is π . Other phase relationships are possible, and as long as the relationship remains constant, we can consider entrainment to take place. However, in experimental studies it has been found that other phase relationships are not as stable as the anti-phase and in-phase ones (R. C. Schmidt & Turvey, 1994). The stability of these relationships is also dependent on tempo, as Scott Kelso and colleagues famously showed in their finger-wiggling studies (Kelso, 1984; Kelso, Holt, Rubin, & Kugler, 1981).

2.2.6 Interpersonal interaction

Nothing in the definition of entrainment suggests that the rhythmic process would need to be produced by a self-aware or even a living thing (c.f. figure 2.2). In fact, the synchronisation and entrainment that Pikovsky et al. write about is mostly a characteristic of any dynamical system, such as clock pendulums or metronomes, as in the original observations of Christian Huygens in the 17th century that gave rise to the field of study of oscillators, pendulums and periodicity. Synchronous group behaviours in animals have also been studied. Joint actions such as swarming, flocking, herding or schooling, depending on the species, have been observed, studied and modelled (Okubo, 1986). A famous example of synchronised pulsations in nature are fireflies flashing in synchrony (Buck, 1938; Buck & Buck, 1968; Hanson, Case, Buck, & Buck, 1971; Strogatz, 2003). The mechanisms of entrainment of the fireflies were measured and modelled; after the initial observations, more controlled studies have been done, using individual fireflies and artificial flashes, looking at the range of their entrainment abilities (e.g. Hanson et al., 1971; Moiseff & Copeland, 2000).

However, I am looking at entrainment from a social cognition perspective, as a mechanism of interpersonal interaction of intentional agents. Therefore, the concept of *agency* is important. I will adopt Albert Bandura's definition: "To be an agent is to intentionally make things happen by one's actions" (Bandura, 2001, 2). I will focus on the entrainment of intentional agents, mostly ignoring the mechanistic entrainment of oscillators (metronomes, pendulums or fireflies), even though they have theoretical value. The intentional, flexible, and social entrainment of humans is likely to be based on such mechanistic entrainment processes at a subpersonal level, on neural oscillations and entrained firing of populations of neurons. Neuronal synchronisation and phase coherence of neural networks are important facets of how the brain works (Fries, 2005). Similarly, phase coherence across brains of people who are interacting with each other is an important neuro-marker of joint action (Lindenberger, Li, Gruber, & Müller, 2009; Sängers, Lindenberger, & Muller, 2011; Sängers, Muller, & Lindenberger, 2012). The neural mechanisms of interaction and joint action have only recently been studied in actual, social settings Hari and Kujala (2009), and they will mostly be outside the scope of this thesis.

Bandura's notion of agency is a response to the information theoretic-

cal view of cognition, brought about by the generation of mind-as-computer psychologists, and a defense of human agents as intentional individuals. In this thesis, I will advocate this view, but emphasise that the intentions are very often social, and that this interaction of independent agents that have social intentions is often messy, but always extremely interesting. In this context, the definitions and criteria for entrainment discussed in the sections above must be seen in this wider context of social agency. One important feature of it is that for every agent that has intentions to make things happen, there are other agents in that social environment who have their own intentions and are able to act upon them, thus generating mutually interconnected dyads or groups of agents. This mutuality and interconnectedness means that the whole of these systems is more than just the sum of their parts. One agent's outputs are inputs to other agents, and vice versa, generating networks with complex, emergent properties (Nowak & Vallacher, 1998; Vallacher & Nowak, 2002).

Often in research this mutuality is not studied as such, as we are so used to testing individuals in isolation. Our theories, methods, measures and analyses are all geared towards this type of testing. Just as the actual, mutual interaction is qualitatively different from the individual performance, studying it is very different and requires novel tools, theories, methods and measures. For example, in the non-mutual approach, the presence of synchronisation can be tested using perturbations. The test subject synchronises with the stimuli, but we do not know if this is just because the periods of the two processes match by chance, so we can perturb the stimuli that we control, make it speed up or slow down, or introduce a phase error, and watch if the subject corrects their behaviour and maintains synchrony. This is a good way of investigating whether two processes are in sync, but the concept of perturbations and error correction change when we move to the dyadic world of mutual adaptation. As the beat is jointly abstracted and redefined on a continuous basis, the ground truth of "correct" beat timing does not exist. As perturbations are the norm and not the exception, *error correction* should be renamed *continuous mutual maintenance*, as it becomes impossible to say what constitutes an *error* and what would *correction* then be.

The new approach to interpersonal interaction and especially entrainment builds on one hand on what has been learned about timing abilities,

spontaneous motor rates, synchronisation accuracy, negative mean asynchrony etc., by testing individuals, and on the other hand what dynamic systems research can tell us about entrainment in general. The challenge is to build a coherent theory on cognition that can take into account the fundamental sociality and mutuality, the emergent nature of group interaction, and then develop a research programme on it.

2.3 Traditional cognitive approach

In the following sections, I am contrasting the embodied, embedded, enactive, social, interactionist etc. views to the “traditional” cognitive approach. It is therefore necessary to briefly explain what this “traditional” view entails. Most authors writing about these “new” approaches use some form of “traditional” or “mainstream” cognitive science as a yardstick to compare their approach with. Necessarily, all these descriptions are simplifications, perhaps even caricatures of mainstream cognitive research, which of course leads to a significant risk of the straw-man fallacy. This means conjuring an image of mainstream cognitive research with obvious shortcomings only to show that your novel approach fixes those problems, leading you to falsely conclude that your approach is better than the traditional one.

The risk for this fallacy is not too high, though, as the main factors that differentiate the traditional view from the embodied, situated and social approaches are not considered shortcomings by mainstream cognition scholars, but rather as design features of cognitive science. There is no major argument in terms of what these differentiating features are. Rather, the debate is on which characteristics of the human mind are considered as most important, as this then determines which model of the mind is more relevant.

The traditional stance on these differentiating factors could be summarised as being *disembodied*, *individualistic*, *representation-based* and *ego-centric*. The alternative strands counter all or at least some of these tenets. I will briefly explain what these four mean, and discuss the alternatives in more detail in the remainder of this chapter. In the context of music cognition research, Marc Leman (2008) has discussed the finer details of the various philosophical and epistemological currents of the field in his recent book. Lawrence Shapiro offers a brief characterisation of the traditional view

in his book on embodied cognition. According to him, the ontological stance of the traditional cognitive science is that “cognition involves algorithmic processes over symbolic representations” (Shapiro, 2011, 2).

The mainstream is decidedly *disembodied* in that it focuses on processes of the mind, and sees the body merely as one source of sensory input and an actuator of motor output. The focus is on the internal processes of the mind. Even the study of social cognition is traditionally *individualistic*, as individuals and their reactions to social stimuli, such as faces or observed social acts are tested. According to the traditional view, the mind operates on abstract, mental *representations* of reality, as an information processing system much like a computer. This view is especially strong in the early cognitive science and artificial intelligence, where computer software and cognitive models were developed in parallel. All these wrap together into an *egocentric* view of the mind, and the view that arises of the mind in its bodily, social, environmental context is in general very different from the one advocated in the embodied etc. camp.

The advantages of the mainstream and alternative views on cognition have of course been extensively debated. One such debate between the mainstream and the (radically) embodied approaches is documented in a special issue of the European Journal of Social Psychology (K. Marsh, Johnston, Richardson, & Schmidt, 2009; K. L. Marsh, Johnston, Richardson, & Schmidt, 2009; Sebanz & Knoblich, 2009; Vallacher & Jackson, 2009). Especially the embodied approach has gained enough traction to invite mainstream cognitivists to assess its claims and contributions, its merits and shortcomings. In the next sections, I will discuss the embodied and social approaches and their benefits in more detail. Because the embodied and social cognitive approaches have somewhat separate historical and theoretic backgrounds and aims, I will discuss these two in separate sections, even though they later converge to the social, embodied approach to cognition that I defend in this thesis.

2.4 Embodied cognition

2.4.1 Embodied approach to cognition

One could argue that the embodied approach to cognition is as old as modern, experimental psychology. Cartesian dualism, the separation of body

and mind, had been criticised already by William James and John Dewey, the late 19th century psychologists, who started the experimental tradition that still continues. Dewey, for instance, argued against the dualistic views that separated the mind and the body, the central and peripheral nervous systems, or the stimulus from the response (1896). He suggested the concept of the “reflex arc” to serve as the unifying principle that would tie together the various empirical findings on sensorimotor processes that started to crop up from different experimental psychology laboratories around the world. William James’s motor theories of perception and of the “ideomotor principles” also would fit the embodied cognition bill (James, 1890).

Even though these early psychologists had already identified the central role of the body in perception, embodied cognition research properly started to take shape almost a hundred years later, as a response to the “mind-centric” cognitive science and its view of the mind as an information processor. In general, embodied cognition refers to considering the body as part and parcel of the cognitive apparatus, more centrally than just as one source of inputs to cognition. A key distinction between the “cognitivist” and “embodied” accounts of how the mind works, is their approach to representations. I will return to this point in a later section.

Maurice Merleau-Ponty (2002 (1962)) is often cited as a key theorist in this respect, viewing the body both as a physical structure and a phenomenological, experiential structure; something that is both inside and outside the mind (Varela, Thompson, & Rosch, 1991). Embodied cognition is far from being a unified theory. It comes in many flavours, and it is contributed to by many scientific disciplines.

Music cognition is a natural context for embodied theories, and an embodied approach to music cognition has been proposed by Marc Leman (2008). He writes about embodiment and mediation, as he sees the human body as a mediator between the mind (ideas, beliefs, intentions, significations etc.) and matter (physical energy of the sound). This mediation goes both ways, as he puts it: “The embodied music cognition approach assumes that the (musical) mind results from this embodied interaction with music” (Leman, 2008, xiii).

Embodied approach thus seems to meet a longstanding demand, and there has been substantial systematic research in the field since the 1980’s. Despite this, the field is still somewhat undefined, and the recent increase in

embodied research has mostly only added to the confusion. At the same time, the concept of embodiment is very versatile and has been applied to many different domains of human behaviour (J. I. Davis et al., 2012). Naturally there have also been various attempts to summarise what the embodied approach is about. I will now discuss three of them.

2.4.2 One, three or six types of embodied cognition?

Many researchers write about embodied cognition these days, and also in music research, there is an increased interest in it, mainly related to music and movement. The term *embodied cognition* is often used loosely, almost as a buzzword, in context of many very different kinds of studies. Andrew Wilson and Sabrina Golonka (2013) discuss the definition of embodiment: there is a “light” version of embodiment, which simply means that “states of the body influence the states of the mind” (Wilson & Golonka, 2013, 1). In their view, this is not what the term should mean. Instead, they map out four steps of analysis needed in a study of embodied cognition. Following these steps, the outcome is radically different from traditional cognitive analyses. Rather than just extending the traditional cognitive analysis by allowing the body to influence the mind states, one achieves a dynamic analysis of a cognitive system that includes both the body and the mind. This, they argue, helps to get rid of the idea of states of mind altogether.

The four-step analysis that Wilson & Golonka propose starts from task analysis from a first person perspective, followed by an analysis of the resources (of the mind, body and environment) that are available for the agent. At the third step, task demands and task resources are put together into a model that allows the agent to solve the problem she is facing. At the fourth and final step, the agent’s performance in solving the problem is tested in such a way that we can confirm whether the model is a good fit for the behaviour. Wilson & Golonka’s model, in particular steps 1–3, is similar to that of Hutchins (1995): it is essentially about redefining the boundaries of the cognitive system, so that it includes the body and aspects of the environment, as well as the individual mind (see chapter 2.4.4).

Lawrence Shapiro (e.g. 2007, 2011) has written about how embodied cognition relates to the traditional approach to cognition. He boils the various embodied approaches down to three themes: “Conceptualisation”, “Replacement”, and “Constitution”. I will discuss these in more detail below. Shapiro

also argues that embodied cognition is not yet a fully-formed *scientific theory* like traditional cognitive science can be considered to be, but rather a more loosely defined *research programme*. In his view, the only common denominator between the various strands of research into embodied cognition is that they are all interested in increasing the role of body in the understanding of how the mind works (cf. Wilson & Golonka’s light version of embodied cognition).

While Shapiro has an “insider view” to embodied cognition, it being his own field of study, Margaret Wilson (2002) has an “outsider view”. She contrasts the embodied approach to the traditional one, just like Shapiro, but she does it from the point of view of traditional cognition, and ends up being more critical. She divides embodied cognition into six aspects, “claims” that embodied cognition makes: “(1) cognition is situated; (2) cognition is time-pressured; (3) we off-load cognitive work onto the environment; (4) the environment is part of the cognitive system; (5) cognition is for action; (6) offline cognition is body based” (M. Wilson, 2002, 625).

These cluster into a few themes. First, there is the relationship between the individual agent and her environment, of which three of these points address directly (1, 3, 4). This is the crux of embodiment, the definition of the cognitive system and whether cognition is confined to the individual mind or whether it extends to the body (embodied), other agents (social), the environment (situated) etc. The other central issue has to do with the time-scale of cognition that Margaret Wilson addresses in points 2 and 6. I find this to be as crucial as the first theme, as it is related to the dynamic nature of cognition and interpersonal interaction.

For both Wilson and Shapiro, an important question is whether the traditional and embodied views of cognition can co-exist. They think that we can learn from both programmes, but many proponents of the “radically embodied” approach to cognition disagree (see e.g. K. L. Marsh et al., 2009; Vallacher & Jackson, 2009), and maintain that embodied cognition is a fundamentally different paradigm that *replaces* rather than *complements* the traditional view. According to the “radical” opinion, if cognition is embodied, it is dynamic, and can not be based on mind states, representations and symbolic processes, as traditional cognitive science claims.

In the next section, I will discuss these various definitions and divisions of embodied cognition in order to reach a working definition of what em-

bodiment and embodied cognition mean, and how they relate to traditional cognitive science. I will structure the discussion under four themes: body movement, situatedness of cognition, timescales of cognition, and representations.

2.4.3 Body movement

According to the Gibsonian embodied view, cognition is *for* action. As M. Wilson notes, embodied cognition makes another claim that goes even further: even offline cognition is body-based, or grounded in bodily metaphors. Of Shapiro’s three types of embodied cognition, “Constitution” is closest to what M. Wilson refers to. Constitution refers to the approach where the body serves as the constitution for cognition, so that the body (or the outside world in general) has not just a causal role on cognition, but rather a constitutive one. The body is an integral part of the mind, not just something external that has an influence on it. This definition of the cognitive system is nicely compatible with evolutionary psychology, and it not only suggests that cognition is *for* action but it also sees cognition *as* action itself.

Support for this theory comes in the form of behavioural and neuroscientific studies on action-perception cycles. Studies on mirror neurons demonstrated direct links between the visual and motor systems, as the same neurons fire both when executing and observing hand movements (Gallese, 2000; Prinz, 1997; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). In short, we understand the actions of others by “experiencing” them in our own mind. This “common coding” (Brass, Ruby, & Spengler, 2009) suggests a model of the mind that is fundamentally different from the traditional, modular views (Fodor, 1983).

Not all studies on body movements are embodied cognition studies, however. It is possible and quite common to study the kinematics of body movements without considering the mind at all, or studying the two in parallel. The important factor is to consider the two together, as parts of the same system.

2.4.4 Situated cognition: mind and the environment

In the redefinition of the cognitive system, the environment needs to be considered, as well. It is not just a backdrop, a context, but just as the

body, a constituent part of the cognitive system, especially as it is often a social one, with other agents in it. Edwin Hutchins (1995) is a key proponent of this approach. Hutchins studied operating and navigating various US navy ships, and suggested that in order to understand what takes place, the whole crew and all the instruments on board, as well as the environment and the port staff should be included in the cognitive model. The ship has its “sensory organs”, the various radios, gauges and radars, and “motor effectors”, its engines and rudders allowing it to move around and influence its environment. These are operated by different crew members, and while the captain has overall control, not all information passes through him or her. This is perhaps an extreme example of situated cognition, but similar ideas have been proposed at smaller scales, from Gibson’s affordances (1977) to ideas to incorporate musical instruments into the cognitive system of a musician (Leman, 2008).

Margaret Wilson sees the strong claim of situatedness—that all cognition is situated—a very controversial one, while she finds more support for the weaker claim that agents use the environment (e.g. artefacts) to offload some cognitive work to it (M. Wilson, 2002). Examples of trying to overcome the limitations of our cognitive system by offloading are writing notes, using fingers or an abacus to help in counting, drawings to remember events etc. This weaker form of the claim is close to Shapiro’s first type of embodied cognition, *Conceptualisation*, that in turn is related to Gibson’s affordances: the body constraints the way in which we conceptualise the environment. In other words, we would understand the surrounding world in a different way if our bodies were different. This way of thinking is somewhere in between the “radical” embodied views and the more narrow view that Wilson acknowledged.

Margaret Wilson also proposes that even though cognition might be situated/distributed, this is not a strong argument for always studying it in that situated/distributed setting. She gives the example of hydrogen atoms, and says that although in nature, hydrogen atoms appear in interaction with other elements (in chemical reactions, forming molecules etc.), we have learned more about them by studying them individually, isolated from those interactions and compounds. Such isolated and non-distributed study would in her view help us find the unchanging, constant principles of cognition instead of just observing the endless parade of individual cases and examples of

how this cognition appears in the distributed reality. I find this a bad analogy. It suggests that the unchanging principles in social psychology are as fundamental and universal as the structure of atoms. This assumption is of course in line with the idea of psychology as a natural science, a mainstream conviction of experimental psychology that has been ardently challenged by since its conception. William James and Wilhelm Wundt, often credited as the fathers of modern psychology, had reservations about the potential scope of scientific methods in psychology, and they acknowledged the cultural and social layers in behaviour and mind much better than Wilson seems to (James, 1890; Wundt, 1904).

There is individual and situational variability in cognition, whether you acknowledge it or not. Acknowledging it gives the researcher a chance of understanding how these constraints work, ignoring them leaves the researcher partially blind. Crucially, a psychology laboratory is not a “non-environment” that magically removes the environmental and situational effects from the experiment. A lab where an isolated individual interacts with a pre-programmed computer is a very specific social environment that does allow for making the environmental and social effects relatively *constant* across participants, but it does not *remove* their effect. Furthermore, it is a very peculiar environment that biases behaviour and cognitive processing in ways that might make the results of these studies difficult to generalise to the “real world”.

The mind is also much more complex than an atom, and thus the reactions of the individual mind with others are not as predictable as the interactions of atoms are. These dynamic, interpersonal interactions need to be studied *in vivo*, as will be discussed in the next section. Wilson is of course right in that we can learn about the workings of the mind also in these more controlled settings, but it will not yield us a more clear view of the operating principles, just a more constrained one.

2.4.5 Timescales of cognition: the dynamic view

According to the embodied view, cognition is “time-pressured”, or it happens in real time, on-line. This is an important aspect in embodied cognition, as it means that agents do not have time to construct and operate on complex mental representations. And according to the embodied view, they do not have to, as the environment is available for them all the time, through their

continuous interaction with it.

M. Wilson argues that human cognition also has an abstract, non-hurried side; an armchair mode, if you will, and thus this argument is not sufficient to discard the traditional cognitive approach. Indeed, given the range of human cognition (memories, complex plans, large structures in language and music, etc.) it is difficult to see that human cognition could operate at only one timescale. There are models suggesting we have two different systems of thinking: the fast and intuitive one for the quick reactions, and the slow, accurate one for more measured decisions (Kahneman, 2011). Similarly, in social interaction, entrainment and coordination are fast and occur ostensibly “without thinking”. In this mode, things happen too fast for there to be time to build complex representations and representations of representations, internal simulations and models and then iteratively modifying these in the light of incoming sensory information.

However, in addition to this automatic coordination, there seems to be another level of social interaction or social understanding. The processes involved with affiliation and empathy are slower and fit easier with the traditional social cognition theories, such as the theory of mind. These two different systems might be subserved by different brain structures, but there is currently very little neuroscientific literature on actual interaction, across these various timescales, but work on this is ongoing (see the concluding chapter for some examples and discussion on Vittorio Gallese’s framework for social cognition that incorporates these different time scales into one model).

Just because a cognitive process is not too time-pressured to involve complex representations and simulations does not of course mean that they exist. Even complex outcomes can result from simple dynamics, and while we are more used to studying empathy or affiliation as states, it does not mean that they could not be studied in a dynamic framework.

Wilson & Golonka were right in emphasising the dynamic aspect as a cornerstone of embodied cognition. In the context of musical interaction, for example in cooperative tapping, the dynamic interpretation is that partners mutually adapt to each other continuously and their synchronicity emerges from the entrainment of underlying oscillations. The traditional cognitivist version of this is that every beat is governed by a new iteration of a linear algorithm that adjusts the timing of a motor command. This algorithm

has a very short memory (usually just one beat long) and period and phase have no meaning, as there are no oscillating processes⁵. According to the definitions in section 2.2, these linear algorithms synchronise, they do not entrain.

In Shapiro’s three-type model of embodied cognition, the dynamic approach is an example of *Replacement*, where representations are replaced by continuous interaction between the agent and its environment. While Shapiro refers to Ellen Thelen’s work with infants and learning to walk (Thelen & Smith, 1994), even he does not seem to give full credit to the fundamental change that the dynamic approach brings about compared to the traditional static view. I will return to this topic in a later section in this chapter (2.6, on page 60) when discussing interactionism.

2.4.6 Representations

The issue of representations has been constantly debated in cognitive sciences over the years. Traditional cognitive science is based on the idea that cognition is about operations on mental representations of reality. Margaret Wilson summarises: “In short, our ability to form mental representations about things that are remote in time and space, which is arguably the sine qua non of human thought, in principle cannot yield to a situated cognition analysis.” (Wilson, 2002, 626.). Shapiro replies succinctly: “Why bother with a representation of the world if the world is right there in front of you?” (2007, 340). He asks whether the representation itself needs to be represented for it to be useful for cognition, which would of course generate an infinite regress of nested representations. As an explanation of how the mind works, the concept of mental representations generates as many problems as it solves, just like the older metaphor of a homunculus would generate a similar, infinitely regressing nested nightmare where the “seat of consciousness” would reside in the homunculus within the homunculus *ad infinitum*.

The embodied, or *enactive* approach thus eliminates representations completely, and sees cognition as actions and interactions (Varela et al., 1991). Mental representations of the environment are not needed as it is available to the mind all the time. Importantly, meanings of things and actions are context-dependent and emerge from this interaction; they are ecologically gen-

⁵Although sometimes components of the linear equation are referred to as phase and period.

erated, based on affordances, and are often subconscious rather than conscious (K. R. Gibson, 1993; K. R. Gibson & Ingold, 1993; K. L. Marsh et al., 2009).

Language is a central cognitive function, and current cognitive models of it are traditional information processing ones, strongly based on representations. In particular, the issue of *meaning* has been difficult to explain from the embodied perspective. Andrew Wilson and Sabrina Golonka argue that meaning is problematic only from the third person perspective, and from the first person perspective of the agent, the task of responding verbally to a partner is identical to the task of catching a baseball: using the available resources to produce an appropriate response, detecting a relevant structure and using it. They argue that in both cases, the agent demonstrates understanding of the meaning of the information (physical or verbal) in their surroundings by producing an appropriate response (catching a ball / responding to a greeting).

Information is information, Wilson & Golonka argue, and that language can be seen as being about achieving communicative goals, just as easily as performing a complex motor act could be seen as achieving its goal, such as catching a ball. They talk about “energy arrays”, patterns that are observable results of the underlying dynamics that identify real-world events. These patterns are “perceptual information” for the agent, and Wilson & Golonka see that at this level, all physical events from seeing the ball to hearing an utterance are equivalent, and a theory of cognition that is grounded on this level does not need higher level representations to work. In the case of the flying ball, its motion, visually perceived, has a certain structure, and this can be used in learning to catch it. In the case of the utterance, it is the pattern of energy in the sound wave that is perceived. Both patterns carry a meaning that is associated with the pattern by convention.

The dynamic models of language are currently applied for explaining language production (e.g. Cummins, 2005, 2009) and how meaning arises when language is used, either through mental simulation (Barsalou, 1999), or by convention (Chemero, 2009). M. Wilson’s challenge about language and meaning that is “remote in time and place” from the embodied situation is difficult to account for using the dynamic, enactive perspective, partly because these offline language abilities are not seen as being so important.

In musical interaction, it is attractive to consider all prior musical knowl-

edge (about the musical style, the context, the partner, the piece, it's structure, typical melodic patterns, riffs, metrical schemata etc.) as mental representations, that then can get aligned in the joint action (Garrod & Pickering, 2009; Tomasello & Carpenter, 2007). In appendix A, the section on metrical schemata describes a pilot experiment that was conducted to investigate the role of matching and mixed metrical schemata in dyadic entrainment.

2.4.7 Embodied vs. traditional cognition

The so-called traditional cognitive science is more clearly defined than its younger challenger, embodied cognition. Traditional cognitive science has a canonical subject matter, set ontological basis and methodology, and thus it can be characterised as a scientific theory. In contrast, there is more uncertainty about what embodied cognition actually is, and thus it has been described as a research programme instead (Shapiro, 2007, 2011). Embodied cognition, in its various flavours, has challenged some of the main tenets of traditional cognitive science, such as focusing on the individual mind; the reliance on representations; the assumption that there are fundamental and immutable basic cognitive processes that are best studied in laboratory using a set of methods that measure mind states and the algorithmic progress from one state to the next.

The traditional and embodied cognitivists have an important difference in focus: the embodied approach is based on dynamics, it tries to explain the on-line nature of cognition; how we can react to the world so quickly; how we can predict events and respond to them appropriately and automatically. The traditional approach is based on more complete, slower off-line processing, trying to account for the complexity of these systems not via dynamics and emergence, but via a complex set of representations, modular structures and algorithms.

To put it in musical terms, embodied cognition is trying to explain *improvisation* while traditional cognition is trying to explain *composition*. These are two distinct goals that perhaps require different approaches. Wilson & Golonka make a very important point that Margaret Wilson perhaps has missed due to the epistemological blind spot that traditional cognitive research often generates: behaviour and cognitive processes can be looked at from different perspectives. Traditional cognitive science typically takes the third person perspective, looking at cognition from an outsider perspective.

From the first person perspective things can look very different: the environment is a set of affordances; a series of challenges to which appropriate solutions and responses are learned. From this perspective, the agent interfaces with the environment by engaging with whatever relevant traces of underlying dynamics the environment has to offer. Sometimes these traces are linguistic, sometimes musical, sometimes based on light changes of the circadian cycle etc. The conviction of the embodied cognitivist is that from the interactions of relatively simple dynamic processes, complexity emerges.

In the end, I agree with Shapiro (and perhaps Margaret Wilson) in that both traditional and embodied approaches can be useful. Depending on what level of the mind you want to study, different approaches may have the most appropriate tools and theories. Trying to explain one of the most complex systems in the universe leads inevitably to a situation where there are multiple approaches that are to some degree mutually exclusive. Anthony Chemero puts it well in the introduction to his book on radically embodied cognitive science: “In fact, nearly everyone working in cognitive science is working on an approach that someone else has shown to be hopeless, usually by an argument that is more or less purely philosophical.” (Chemero, 2009, 3). And most people will continue working on their adopted approaches regardless, as they see that they can learn most about their favourite aspect of the human mind using their tools. Also, as Shapiro puts it (2011), while there is an overlap between the subject matter of traditional cognitive science and embodied cognition, the latter also expands the scope of research to areas that might not be interesting at all for traditional cognitive science, such as babies learning to walk or catching balls. Again, the approaches are not necessarily at odds with each other, but complement each other.

2.4.8 Embodied cognition in music

Music is a rich domain for embodied cognition research, with many relevant aspects, as I will discuss in section 2.8. Music performance requires very highly developed motor skills. But in addition to this, also Merleau-Ponty’s “inner” version of movement–movement as a metaphor–is well-established in music. We perceive music as flowing, galloping, rocking and rolling, going higher or lower, faster or slower etc.

A connection between physical movement and metaphorical movement has been discussed for example in the context of the final ritard, the slowing

down at the end of the musical piece that has been compared to the slowing down of a physical movement, such as a runner slowing to a stop (Friberg & Sundberg, 1999). Friberg's and Sundberg's was a modelling study, and this issue is discussed from a more philosophical point of view by Henkjan Honing (Honing, 2003). An interesting way to approach this nexus of metaphorical and physical movement experimentally was employed by Marc Leman and colleagues in their study on the Chinese stringed instrument Guqin (Leman, Desmet, Styns, Van Noorden, & Moelants, 2009). This style of music is melodic and rich in glissandi and ornamentation. In this study, they recorded the movements of a performer using an optical motion capture system, and then compared this data to data collected from listeners who had moved their arms in response to the music. They found that the velocity profiles of the Western listeners' arm movements were correlated to each other, and also to the shoulder movements of the performer.

These studies, as most having to do with the virtual movement, look at velocity, which translates to the tempo of the music. In the guqin study, the connection is between the shoulder movements of the performer and the hand movements generated by the listeners. The shoulder movements can be interpreted as being both sound generating movements, as well as ancillary or expressive movements (Davidson & Correia, 2002; Thompson, 2012; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). Thus the participants, none of them familiar with guqin music, were not simply tracking the physical gestures needed for sound generation, and as such did not just follow some acoustic dimension (pitch or intensity, for example), but rather a combination of the acoustic features and the expressive intentionality of the performer. Using terminology from Wilson & Golonka (2013), the underlying dynamics of the performance are encoded into acoustic signal, and this "energy array" is perceptual information for the listener, who in turn entrains to it and produces a movement that traces some aspects of the original system.

Metre is a topic that fits very naturally within the embodied cognition programme. It can be seen as an abstract phenomenon, but it has a very practical, corporeal implementation in music-related movement. Examples of such studies are the African dance -study discussed in chapter 8, and the paper on spontaneous movement to music by (Toiviainen, Luck, & Thompson, 2010).

Another example is the series of studies on “bouncing babies” by Jessica Phillips-Silver and Laurel Trainor (Phillips-Silver & Trainor, 2005, 2007). In these studies, the developmental aspects of metre perception were studied. The participants were 7 months old, and they were bounced to an ambiguous rhythm, either in duple or triple time. After that, they measured which version of the ambiguous rhythm babies liked better, showing that they preferred the one they had been bounced to. Babies showed this preference regardless of whether they had been blindfolded during the bouncing or not, thus excluding the possibility that the effect is due to different visual conditions. In the control group where babies had just passively observed the experimenter bouncing but had not bounced themselves, there were no differences in their preferences of the two variants of the stimulus, indicating that embodied experience and activation of the vestibular system in the ear is required for the effect.

In the adult version of the experiment, participants bounced by themselves by bending their knees in time with the unaccented, ambiguous stimuli, either in duple or triple metre. After this training stage, they heard the two accented versions of the stimuli, and had to select which of them was the one they had heard in the training stage. Just as the babies indicated their preference for the metre they had been bounced to, the adults more often chose the version they had moved to earlier.

In terms of the more basic theories of timing, temporal tracking and beat induction, there are many families of theories. Even though they all deal with sensori-motor aspects, the nexus of rhythm perception and production, some are classic cognitivist theories (Desain & Honing, 1999; Mates, 1994; Vorberg & Schulze, 2002; Wing & Kristofferson, 1973), while others are based on connectionist approach (Large & Kolen, 1994; McAuley, 1995). There have also been some attempts to develop a more embodied theory of beat induction, by Neil P. McAngus Todd and his colleagues (Todd, Lee, & O’Boyle, 2002; Todd, O’Boyle, & Lee, 1999). Todd et al. try to model rhythm perception and production, starting from the neurobiological basis. Their model contains a model of the ear for the input, and models of the motor effectors as the output. Todd also suggests that the balance sense in the ear (the vestibular system) also plays a role in rhythm perception, a view that would indeed fit well with the embodied rhythm studies of e.g. Phillips-Silver, Trainor, and Toiviainen, discussed above.

This theory about the role of vestibular system in rhythm perception was put to a test by Laurel Trainor and her co-authors (Trainor, Gao, Lei, Lehtovaara, & Harris, 2009). Their experiment combined the ambiguous (duple or triple) rhythm patterns from the earlier Phillips-Silver & Trainor studies with direct, galvanic stimulation of the vestibular system. A sensation of lateral movement was induced by delivering a small, time-varying electric current to an electrode attached to the mastoid process (a bone behind the ear). This galvanic stimulus was synchronised to the auditory one, generating movement sensations in either duple or triple time. It was found that direct vestibular stimulation without actual body movement worked in a similar way as moving to the rhythms, biasing the identification of the ambiguous rhythms as duple or triple.

All these different families of models have advantages in different aspects of rhythm perception and production, which in itself is a very rich and multi-dimensional phenomenon. Comparing the models and theories systematically to each other is not simple, as finding objective criteria for their goodness is not easy. It might not be necessary, either, as these models are not necessarily mutually exclusive. Therefore the pragmatic approach would be to use models, theories or frameworks that seem to suit the current needs best.

2.4.9 Radically embodied social psychology

Social psychology is traditionally a psychological inquiry of the individual in social contexts. Thus traditional social psychology is “compatible” with and similar to the traditional cognitive science in all the major ontological convictions. The criticisms that embodied and situated approaches have made to traditional cognitive science also hold for social cognition and social psychology. The case for adopting situated and embodied approaches to the study of social psychology has also been made, for example by Kerry Marsh and her colleagues. They suggest a radical departure from the “cognitive constructionism” (K. L. Marsh et al., 2009, 1217). In their agenda-setting paper, they make four suggestions for social psychologists, based on re-thinking the role of the environment and social interaction in social psychology.

Their suggestions are the same as those discussed above: 1) *actions* should be studied as such and not as reflections of mind states (embodiment, or Shapiro’s *Constitution*); 2) actions and the relationships that an

individual and the environment have are *dynamic* (as Wilson and Golonka also emphasise). As Marsh et al. have a social point of view in their paper, their concrete suggestion is very relevant for this thesis: social psychology should study how behaviour unfolds in time. Concretely they recommend studying the coordination of actions by means of looking at interpersonal synchronisation. 3) Marsh et al. also recommend that researchers should focus on *joint participation* in goal-directed action, and 4) as the environment is part of the dynamic system, behaviours should be studied in their *natural settings*. Marsh and her colleagues have been “walking the walk” and used many innovative experimental designs to study interactional dynamics and environmental affordances (see chapter 3).

Marsh et al. call their approach “radically embodied” (see also Chemero, 2009), the radical part being that they want to discard mental representations altogether, and also they are of the opinion that the embodied paradigm replaces rather than complements the traditional view of cognition. Sebanz & Knoblich commented this article in the same issue (Sebanz & Knoblich, 2009), warning that adopting the “radical”, ecological view might lead to theoretical inconsistencies. They find it problematic that the ecological view does not address concepts such as “stereotypes, motivations, intentions, or emotions” that are the bread and butter of traditional social psychological theories (Sebanz & Knoblich, 2009, 1231).

While Sebanz and Knoblich agree with the aims that Marsh et al. have, they promote a more careful, integrative approach that would bring together the existing understanding of social processes, and add to that by shifting focus to the issues that Marsh et al. suggest (joint action, dynamics of social environment, interpersonal coordination). In another response, Vallacher & Jackson (2009) take an even more extreme view and insist that Marsh et al. are on the right track but not radical enough. They argue that the “old” social psychology is wrong altogether, as it does not take into account the emergent, non-linear nature of behaviour and coordination. In their view, dynamical systems are recursive and thus all levels of cognition can be explained with hierarchical models of dynamic systems.

2.4.10 Summary

Embodied cognition is not (yet) a well-defined theory or a model of how the mind works. Rather, it is a general philosophical approach, or at best a research programme that emphasises the central role that body and action hold in cognition. The essential feature is that rather than seeing behaviour as a consequence or a reflection of an internal mind state, cognition is seen to subserve action. Embodied approach means that actions, the body, and the environment are integrally incorporated into the dynamic system, not just considered as contextual features that have some contribution or influence to the abstract, internal, representation-based cognitive processing.

The issue of representations, and the role that the internal, disembodied processing might have divides the scholars of embodied cognition. Marsh, Vallacher, A. Wilson, Golonka and other proponents of the radical flavour of embodiment disregard them completely, while Leman for example sees the body as a mediator between the physical and phenomenological worlds, much like Merleau-Ponty did. Sebanz & Knoblich are also in favour of an integrative view, while M. Wilson speaks strongly for “disembodied” cognition having an important role.

The embodied approach fits very naturally in music research, and on the other hand music is a very suitable domain for embodied research. Music foregrounds the interpersonal entrainment that the “radical embodied social psychologists” study as an index of coordination and cooperation. The situated and distributed aspects of cognition (a group of people and their instruments as the cognitive system) are clearly evident and available for experimental manipulation. Music also combines the abstract and phenomenal with the concrete and physical, and both these to the motivational-emotional processes.

2.5 Social cognition

Humans are a “hypersocial” species, and “wired” to cooperate (Hrdy, 2009, 4). Therefore a comprehensive understanding of how our minds work, requires understanding how this sociality works. And just as discussed above in the embodied cognition section, sociality should not just be an external context that has a causal effect on behaviour, rather, it needs to be integrated into the system. In this section, I will present models of social

cognition that support the idea that in social interaction, the whole is larger than the sum of its parts ($1 + 1 > 2$). In interaction, the behavioural output of one agent becomes the sensory input of the other and vice versa, creating a bi-directional feedback loop and non-linear effects emerge.

2.5.1 Historical approaches to sociality

While the importance of social context has been acknowledged in psychology since the dawn of experimental psychology, even the early social psychology was based on studying the individual (Allport, 1924 (reprint 1994)). However, George Herbert Mead “promoted” the social processes from being just a backdrop for individual processes into having a fundamental role in the continuous process of defining “self”. Mead’s student Herbert Blumer coined the term “symbolic interactionism” to describe this approach.

Blumer argues that the meanings that things have, determine how people act towards them (cf. Gibson’s affordance theory), and these meanings are derived from social interaction and group life. Also, these meanings are handled and modified in an interpretative process. The idea thus is that different groups and at different times can give things different meanings, and the individual interprets the meanings “live”, in the interaction. Therefore, social life should be studied as action, and not excluded from the experiments as a static background or a medium where interactions take place. Blumer states that joint action is the result of fitting our actions together as a joint production. This joint action is not just a sum of the individual actions, however, but rather an entity with a life of its own. (Blumer, 1969; Manning & Smith, 2010).

Even though symbolic interactionism was formulated as a response to behaviourist psychology rather than cognitive psychology (as these theories predate the cognitive revolution), many of its criticisms are relevant for experimental cognitive psychology, as well. For example, Blumer argued that no amount of methodological rigour can replace a deep understanding of the phenomena, which is often acquired through qualitative studies, observation etc. Symbolic interactionism, much like the current dynamic and embodied approach emphasise the dynamic nature of the world, in contrast to the static approach in both behaviourism and traditional cognitive science. The importance of social interaction has been emphasised by a number of scholars since Mead and Blumer. It is related to the situated and embodied

approaches discussed in the previous section, in that it redefines what should be part of the cognitive system (Hutchins, 1995; K. L. Marsh et al., 2009; Semin & Cacioppo, 2008).

2.5.2 Modern social cognitive approaches

Symbolic interactionism was a sociological approach that was very accurate in identifying concurrent shortcomings of the studies of human nature and sociality, but the proposed remedies were not very helpful for psychology or the cognitive science that soon followed. Phenomenological and qualitative, descriptive approaches did not fit the “scientific” epistemology of then emergent cognitive science. Only later did cognitive psychology and cognitive neuroscience mature enough for the embodied and social approaches to gain ground.

The field of social cognition contains many interlinking strands, with more cognitivist theory of mind -based approaches in one corner, embodied, dynamic theories in other, with social neuroscience emerging somewhere in between. The topic of interaction has been studied in the context of conversation and interaction in general (Burgoon, Stern, & Dillman, 1995; De Jaegher & Di Paolo, 2007). In the following sections I will describe some central features of current social cognition theories, and look in more detail at one model that links social neuroscience and music, and another that attempts to weave together the neural, behavioural and phenomenological levels of intersubjectivity.

2.5.3 Mimicry

We have a tendency to mimic the gestures, movements, speaking style and rate of our conversation partner (Giles et al., 1991). This “chameleon effect” has been demonstrated both in observational studies (Kendon, 1970), as well as in experiments (Chartrand & Bargh, 1999; van Baaren, Holland, Kawakami, & van Knippenberg, 2004). Feelings and expressions of emotions have also been found contagious (Hatfield, Cacioppo, & Rapson, 1994).

For example, in the experiments by Chartrand & Bargh (1999), the participants were interacting with two different confederates of the experimenters, and were found to change their behaviours according to the models (smiling, touching faces, moving feet) provided by their current interactants.

The number of these behaviours increased significantly when the confederates had displayed them. This was true for both the facial expression (smiling) and the behavioural measures (face rubbing and foot shaking). Furthermore, they observed that there was no difference in mimicry between smiling and not smiling -conditions, suggesting that liking the other person is not a necessary condition and perhaps at least in these “minimal” conditions not even a mediating factor. The authors refer to this finding as evidence that the chameleon effect is not “goal-dependent”—this is with regards to the previous studies (e.g. LaFrance, 1979) that suggested that the origin of this effect is in trying to increase interpersonal liking or affection.

In the second experiment, Chartrand and Bargh investigated the adaptive functions that this chameleon effect might have. From prior research, they point to theories of humans’ powerful need to belong. Thus, being similar to others around us, mirroring mannerisms etc. would be natural ways to smooth the interactions and display our willingness to belong. In motor mimicry research (Bavelas, Black, Lemery, & Mullett, 1986) the authors note that showing the wince of pain on one’s face upon seeing the other hurt themselves, is not done for intrapersonal reasons only. For example, that gesture or wince is not an external manifestation of an intrapersonal cognitive process such as role taking, or an expression of a “vicarious emotion”, but rather an important communication tool that relays the message “I am like you” or “I feel as you do” to the other person. (Chartrand & Bargh, 1999, 901.)

The third experiment reported by Chartrand & Bargh was looking at individual differences in non-conscious mimicry. The authors focused on the cognitive component in empathy, perspective taking, measured with Interpersonal Reactivity Index (IRI) by M. H. Davis (1980). It turned out that those scoring high in the perspective taking scale mimicked the confederates more in the experiment. However, there was no correlation between the emotional dimension of empathy and mimicry, which led Chartrand & Bargh to conclude that the chameleon effect is modulated by some individual differences, but not others.

Van Baaren et al. used the same experimental method as Chartrand & Bargh and also performed three experiments (van Baaren et al., 2004). In the first one, they used the pencil-dropping test (Macrae & Johnston, 1998) and a mood scale as outcome measures and compared groups where the

confederate had mimicked the behaviour of the participant or not. Of the mimicked group, 100% of participants picked up the pencils, whereas only a third of the non-mimicry group members did. Mood was not interacting and not affecting the pencil pickup rate.

In the second experiment, the setup was the same but this time the pencils were dropped by another experimenter, not the one the tasks had been performed with. Again, the mimicry group picked up more pencils, suggesting that the positive effects are not restricted to the mimicker but to a wider community, or in the words of the authors, “mimicry can produce a diffuse prosocial orientation that transfers to people in general” (van Baaren et al., 2004, 75). This prosocial orientation was then further put to the test in experiment 3, where the outcome measure was the amount of money participants donated after the mimicking manipulation. The mimicked participants donated more than those that had not been mimicked, and again this effect was generalised to giving more money to a person the participants met for the first time after the mimicry manipulation. Therefore, the authors confidently concluded that mimicry increased prosocial behaviour, and that this effect is not restricted to the interaction partner but that mimicking other person’s posture and gestures leads to an increased and general prosocial orientation.

Van Baaren and colleagues conclude with speculation about the adaptive value of mimicry. They suggest three possible mechanisms. First, mimicry might foster safety in groups; second, it might be a tool for acculturation and learning; and third, it serves as a “glue” that holds social groups together. The authors’ findings would indeed seem to support the third mechanism especially, as mimicry not only seems to increase the bond between the one who mimics and the one that is being mimicked, but also more generally have a positive effect on sociality in the group.

2.5.4 Interactional synchrony and affiliation

There are a number of studies demonstrating that entrainment also boosts our pro-social tendencies and has positive emotional impact on us. Michael Hove and Jane Risen showed that tapping in synchrony with the experimenter increased participants’ liking of the experimenter (Hove & Risen, 2009). In their three experiments, they first established a link between the amount of synchrony and positive affiliation ratings, then manipulated this

synchrony to demonstrate a causal direction between the two. Finally, they tested if the same effect would be reached by having the experimenter passively present while the participants synchronises with a metronome. This time no effect on affiliation was found, showing that the interpersonal synchrony is a necessary component for the affiliation effect to occur.

Marcel Zentner and Tuomas Eerola found that even babies show this correlation between entrainment and positive affect. In their study on infants moving to music, they found that the more the babies' movements were coordinated with the music, the more they displayed positive affect, e.g. smiled. Thus this link between being together in time and positive affect seems to be a fundamental one. Music-induced synchrony promotes co-operation (Wiltermuth & Heath, 2009), which in turn is rewarding and pleasurable. This is one of the mechanisms by which music generates its positive effects.

2.5.5 Empathy and mirroring

The neural basis of interpersonal interaction has been studied as well. There has been an increased interest in the social in neuroscience since the serendipitous discovery of mirror neurons (Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti et al., 1996). These are neurons that fire both when observing and performing goal-oriented actions. These neurons were found in Macaque monkeys, but similar mirroring systems have been found also in humans (see Hari and Kujala (2009) for a review).

Istvan Molnar-Szakacs and Katie Overy have proposed a model that combines aspects of mirroring, movement, emotions, and music (Molnar-Szakacs & Overy, 2006; Overy & Molnar-Szakacs, 2009). Their SAME (Shared Affective Motion Experience) model assumes that music is perceived not only as abstract acoustic stimuli, but rather in terms of the intentional motor acts behind the generation of those sounds (see also E. F. Clarke (2005) for an ecological approach to music perception). The temporal cortex and fronto-parietal mirror neuron system then activate as a result of this perception-action equivalence. Anterior insula serves as a conduit between the mirror neuron system (MNS) and the limbic system, leading to the complex affective evaluation of the signal. As this activation occurs both in the performer and the listener, this creates the Shared, Affective Motion Experience.

Motor control in the SAME model is based on hierarchical organisa-

tion (J. Kilner, Friston, & Frith, 2007; J. M. Kilner, Friston, & Frith, 2007). This hierarchy extends from long-term goals (intentions) to shorter term action goals, to the kinematics involved in certain movements, and finally to the neural impulses that contract individual muscles that are involved in the execution of the movements. The mirror neuron system is central in this model, as it is seen as the neural implementation for action understanding. The activation in the MNS during both observation and execution of goal-directed actions such as grasping, provides a plausible “bridge” allowing the intentions of actions to be interpreted just by observing the motor actions themselves. Humans are extremely well-tuned to perceiving biological motion: we are good at reading personal characteristics such as age, gender or amount of effort from point-light displays, which are very reduced representations of biological motion (Brownlow & Dixon, 1997; Johansson, 1973; McArthur & Baron, 1983).

From fMRI studies on observing actions, it has been shown that the MNS activation is stronger when the participant had expertise in producing the actions him/herself (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006). Similarly, practicing to play an instrument changes the brain’s responses to the sound of that instrument (Pantev, Roberts, Schulz, Engelien, & Ross, 2001); music practice induces many other functional and structural changes (Schlaug, Norton, Overy, & Winner, 2005). Different expertise levels thus change how the brain functions, how it processes musical sounds and actions. According to the SAME model, this is reflected in how the different levels of intention and motor control can be perceived. The higher level intentions and goals are more readily perceivable even without significant musical expertise, while between experts on the same instrument, the mirroring can extend down to the level of specific muscle activations.

Molnar-Szakacs and Overy emphasise that the anterior insula has an important role in linking the MNS and the limbic system, thus bringing the affective and empathetic responses into the mix. Empathy here refers mainly to its emotional dimension (M. H. Davis, 1980), although it could arguably be also described as sympathy (Malloch & Trevarthen, 2009; Trevarthen, 2006). This affective side of intersubjectivity is a strong motivator, as we know from everyday experiences, as well as from experiments (Hove & Risen, 2009; McNeill, 1995; Wiltermuth & Heath, 2009).

Essentially, the SAME model attempts to explain how “when we hear music, we hear the presence (or agency) of another person, whose *actions* we can interpret, imitate, and predict.” (emphasis added Overy & Molnar-Szakacs, 2009, 495). This is equivalent to the theory of embodiment: in the sound of music, the traces of actions that generated it are encoded. The addition is, however, that these actions are not simply motor patterns, but socially relevant information, as they codify the relationship of performers and entrain also the listener to being together in time.

Molnar-Szakacs and Overy discuss the implications of their model for music therapy. Their model brings together features of music and musicking with neural mechanisms of intersubjectivity and affect, thus making a plausible explanation for music’s therapeutic efficacy (MacDonald, Kreutz, & Mitchell, 2012; Ruud, 2010).

2.6 Interactionist approach

Although the importance of social and embodied aspects to cognition have widely been recognised, in most studies methods are still from the traditional tool chest. As discussed already in the sections on embodied and social cognition, properly adopting the embodied and social perspectives requires a more fundamental epistemological and methodological shift. In addition to redefining the cognitive system to include the body and the social environment, three central points emerge: 1) the mind is active, 2) the environment is dynamic, and 3) the focus of analysis should be in this interaction, not in the static properties of the mind. Together, these points constitute a shift from the passive, egocentric to active, interactionist point of view.

Depending on the approach, a researcher can have a *first person* or a *third person* point of view to the studied phenomenon. Participatory observation, often utilised in ethnology, ethnomusicology or social anthropology would be an example of the first person view. Here, the researcher becomes a part of the practice or phenomenon she is investigating, in order to learn about it and gain understanding of how it works. In cognitive psychology and neuroscience, researchers usually take a third person point of view; they remain outside of the situation, observing it from afar. These perspectives could also be called “emic” and “etic”, respectively (Berry, Poortinga, & Pandey, 1997; Headland, Pike, & Harris, 1990; Pike, 1967). This distinction goes

deeper in the debate on the cognitive approach, however. The third person approach of cognitive psychology also extends to how the mind is seen. The disembodied, non-social, non-agentic mind has the role of a passive observer in the world. The mind is not actively involved in creating or modifying the environment. Even in the experiments of social cognition, the mind is just an observer of socially relevant stimuli, such as faces or (inter)actions of others. Also, while a participant in a typical tapping experiment is actively generating sounds, they are passive in the sense that they have no effect on the metronome they are tapping with or any other aspect of their environment.

In the so-called *enactive* approach, a combined first and second person perspective would be taken (De Jaegher & Di Paolo, 2007; T. Fuchs & De Jaegher, 2009). To simplify somewhat, this means that the minds that we study are engaged in interaction with each other. They are active agents, not just perceiving some unchanging external environment, but shaping the world surrounding them as they go along.

From the methodological point of view (and to exaggerate somewhat), in the traditional cognitive approach the mind is passive (in the sense that it has no feedback effect on the context). The relationship between the mind and the environment (“the reality”) is thus seen as being static (rather than dynamic), and therefore the timescale of measurement is considered not important. Thus, taking static measurements, such as ratings of stimuli or average asynchronies across whole tapping trials is considered sufficient. The research designs and statistical models are also based on these assumptions: the researcher collects static scores of performance of the participants (or, in this case they perhaps should be called “subjects”) in one or more conditions, and in between conditions, the reality/environment is categorically and in a controlled fashion changed by the experimenter. Alternatively, there are multiple groups of subjects who differ from each other according to a certain categorical measure (such as being musicians or non-musicians). Getting results is then basically a question of either correlating scores or comparing means.

This model works only if the “reality” can be assumed to remain constant. If passive *subjects* in their third person perspective are replaced by active, engaged, first and second person *participants*, the “reality” is in constant flux, and the static measurements no longer work. As Fuchs and De Jaegher put it, “social understanding is not realised by ‘snapshot’ activities of one

individual’s theorising or simulating but arises in the moment-to-moment interaction of two subjects” (Fuchs & De Jaegher, 2009, 466).

In the traditional approach, the idea of averaging is to boost the signal over noise. For example, in synchronisation tapping, a common model predicting tap timing (see e.g. Schulze & Vorberg, 2002) contains error correction parameters (the signal) as well as random noise. By averaging over a number of taps, the signal keeps adding up, while the random noise does not, thus making the signal more visible.

In this model, the subject’s reactions do not change the “reality”, and the assumptions of the model are probably right: more repetitions boost the signal and give a more accurate result. However, contrast this controlled setting with the more naturalistic situation where two active, engaged participants entrain to each other. Here each tap, each repetition of the task takes place in a new environment: every reaction is also an action, and the linear error correction model no longer fits what is going on. The linear model fails not so much because it might not be able to predict the data, but rather it fails at a more fundamental level: it does not fit the dynamic and ever-changing nature of “reality”.

Thus, while the “information-processing approach” (see e.g. Repp, 2005a) is suitable for the one-sided, highly constrained synchronisation studies, it might not be extendable outside that realm (Himberg, 2012). In these more “noisy” and changing environments, solutions based on the dynamic systems (such as adaptive accompanists that find and follow beat using a bank of oscillators, (Toiviainen, 1998)) seem to fare relatively well, while they in turn do not replicate the exact short-term error correction patterns that are observed in the lab. Bruno Repp suggests (2005, 970) that in synchronisation research, the information-processing approach works well in the discrete domain of tapping, while the dynamic approach works better in the domain of continuous data, such as movement data. This is true, although I would add a further constraint: the information-processing approach only works with discrete data when the individual is operating with non-responsive environments, i.e. when reality/environment does not change in response to the subject’s actions.

2.7 Examples of interactionist experiments

The study on motor coordination by Braun, Ortega, and Wolpert (2011) is an example of a clever setup that “automatically” focuses the attention to coordination rather than individual performance. Their study combines motor coordination with game theory. Participants hold handles and move them over a tabletop. The handles are attached to robotic arms, and interconnected via a computer. Participants’ task is to move their handle from one end of the tabletop to the opposite end (along the x-axis). They need to find the path of least resistance across the table by varying their position on the y-axis. The robotic arm resists their movements based on where they are laterally. The resistance and consequently the optimal path depends on the y-coordinate of the other participant, according to rules corresponding to classic coordination games from game theory research. Thus both participants’ actions are mutually dependent on the actions of the other. In all the different games, participants could reach equilibria, or mutually optimal trajectories, by coordinating their movements. The analysis shows that even though the game rules or parameters were changed essentially for every go, participants managed to coordinate and were successful in finding the equilibria.

In another computer-feedback experiment, Auvray, Lenay, and Stewart (2009) tried to establish a setting for minimal interaction. In their experiment, participants moved an avatar on computer screen using a mouse. They were exploring a one-dimensional space (a horizontal line on the screen). They were blindfolded, so they could not see their own or the other participant’s avatars, but through a tactile feedback device they received a tactile stimulation whenever their mouse cursor “met” the avatar of the other participant. The smart trick was that there were three kinds of objects on the line that participant A could meet: the avatar of participant B (which would result in both participants receiving a tactile stimulation), a stationary object (that would stimulate only participant A and would not move), and a “mobile lure” of participant B (that would stimulate participant A but participant B would not notice this encounter). All tactile stimulations were identical all-or-none stimulations.

When operating in this minimal interactive environment, participants were able to distinguish between the stimulations caused by encounters with

the other participant and those caused by the stationary targets or the lures. They preferred and developed strategies to seek these mutual encounters. Participants recognised the “jointness” of these encounters. This result suggests that we are sensitive to mutuality and prefer mutual interactions. This finding is in line with the findings discussed above regarding the affective and prosocial effects of mimicry and entrainment. A similar preference for mutual entrainment over one-sided synchronisation was also consistently found in my tapping studies (see chapter 7 and appendix A).

Computers, robotic arms, force feedback, or tactile stimulators are not necessary for interactionist research. Simple dyadic tapping tasks can be used, and observations of conversations, music duets or music therapy sessions can be conducted in an interactionist fashion. Current technologies also allow the researchers to control and modify complex interaction parameters on the fly, during the experiments. This helps control the interactions occurring in the experiment, and generating different experimental conditions.

2.8 Summary: music as an ideal domain for studying embodied, situated, social cognition

In this chapter, I have discussed embodied, situated, and social approaches to cognition, and presented the interactionist approach that combines many of their features. While these approaches are somewhat different in their foci, they all agree on the shortcomings of traditional cognitive science. According to all these approaches, traditional cognitive science is wrong in limiting cognition to be a function of the isolated, disembodied, individual mind. The body, the environment and other people are only seen as having a causal influence on the mind, rather than being an integral part of the same system with it. These newer approaches also criticise the view that cognition is static, or progresses from one mind state to the next. According to all these approaches, cognition should be seen as dynamic, constantly evolving via an interaction between the mind and the body, with other agents and the environment.

This makes cognition challenging to study. The setting should be as natural as possible, and for studying social processes, there should be immediate, actual interaction. To properly account for the role of the body, participants’ responses should not preferably be limited to button-pressing, but involve

natural movements of the whole body. At the same time, the experimenter must be able to control and systematically vary aspects of the environment, the interaction, and behaviour, as otherwise the analysis can not proceed beyond qualitative observations or descriptive analyses. Inferential statistical analysis, hypothesis testing and comparisons between conditions are part and parcel of cognitive science, and these methods are also central in social, embodied and situated cognition. Thus, on one hand, the situation needs to be as natural and free as possible, so that the behaviour is as “normal” as it can be, and on the other hand, the situation needs to be constrained and controlled so that systematic analysis is possible. Balancing these two needs is a long-standing challenge in all fields of behavioural science.

Music and dance are very well-suited for studying embodied, situated and social cognition, as they have design features that help in finding a balance between naturalness and control. Music and dance are not only very malleable, but constraining and changing the interaction dynamics between performers is a natural dimension of artistic expression. Musicking and dancing are quintessentially social pursuits with a range of different kinds of inter-participant relations. There can be clear leader-follower roles, coordination can be synchronous or sequential, there are call and response patterns, imitation as well as variation etc., etc. The wide range of musical genres and practices, huge flexibility in terms of performance contexts, and readily available theoretical descriptions of their content provide almost endless possibilities for ecologically valid experimental manipulations.

While both music and dance can be notated symbolically, they are obviously embodied and have a strong motor component, and are not as representational as is language. Symbolic representations and algorithmic processes on those symbols have been essential in language studies because they are tools that access the level of *meaning* in language. Music and dance lack this referential meaning, and even when looking at music and dance as forms of communication, they operate on more “floating” types of meaning (I. Cross, 2005). Thus it is easier to move away from representations as basis of cognitive functioning, when studying music and dance than it would be when studying language.

Michael Thaut repeatedly refers to music being the ideal setting for studying cognition in his book *Rhythm, Music, and the Brain* (2008). He states, for example, that: “To put it simply, music may be a language the

brain can read with ease because its temporal-based grammar is fundamental to how the brain processes information” (Thaut, 2008, 79). He argues that rhythm in music entrains the brain in a way that no other stimulus does; it links the auditory and motor areas, as well as cortical and subcortical structures (c.f. the SAME model by Molnar-Szakacs & Overy, 2006).

Behaviourally, music performance requires a wide range of abilities and skills. Much focus is put on the fine motor skills of music performance (e.g. Palmer & Meyer, 2000), and on auditory perception skills, but formal music training also involves continued, focused attention, development of regular routines, matching sensory information and motor inputs, learning to read notation, learning attention to details, memorising pieces of music, entraining with partners, expressing and interpreting emotions and empathy, all of which provide entrees for study (I. Cross, 2008; Schellenberg, 2004).

Many higher order cognitive processes can be indexed using deceptively simple tasks and measures of auditory perception, such as the mismatch negativity (MMN), which is a signal of preattentive auditory discrimination, or the dichotic listening paradigm, which is a simple test based on brain asymmetry (Hugdahl et al., 2009; Näätänen et al., 2011). Both have been used as indexes of a surprisingly wide range of developmental and neurocognitive disorders, ranging from dyslexia and autism to schizophrenia and Alzheimers, and from sleep deprivation to alcohol use.

Music is a human universal, and tightly intertwined to the evolution of abstract thinking, language and the capacity for culture. The potential to transcend cultural boundaries makes music more attractive than language for cross-cultural investigations, yet the subtleties of meaning afford even high-level meaning-making to be investigated in a non-trivial fashion.

Methodologically the shift to a social, embodied approach is challenging, and requires a shift of focus from static to dynamic approach, and from individual to interaction. In the next part of this thesis, I will discuss these methodological issues. In the next chapters, I will review different ways in which coordination and synchronisation have been studied before, paying special attention to methods and measurements. First, I look at studies on detecting entrainment. The criteria for entrainment were discussed earlier, in (2.1), and now the issue of measurements and thresholds is also tackled. After that, I will review previous studies that have looked at dyadic and group entrainment. After this literature review, I will drill down to details

of methodology, taking a look at measurements and statistical tools that can be used in analysing interaction. The main focus will be in circular statistics, a set of methods used for looking at periodic data such as musical rhythms, gait etc. After that, I will present a number of other methods for detecting, visualising and quantifying interaction. Some of the more technical details of these methods can be found in appendices.

In chapters 7 and 8, as well as appendix A I will present empirical studies on musical interaction. While they do not cover the whole spectrum of music and dance experiments, they are examples of social and embodied approach, and hopefully help in illustrating the flexibility of music and dance in experimentation.

To summarise the current chapter: whether we study music as “humanly organised sound” or investigate its effects on a “soundly organised humanity” (Blacking, 1973), it provides us a testbed for experimenting with all layers and levels of social, embodied cognition and behaviour.

Part II

Interaction - prior research and methods

Chapter 3

Detecting synchrony and entrainment

3.1 Introduction

There can be many different goals and research questions that require studying interpersonal entrainment and coordination. Often, as in music and dance studies, the objective is to analyse how well participants are entrained, how closely their movements are synchronised. However, sometimes the researcher is looking for a yes or no -answer, checking whether there is entrainment or not. Detecting entrainment is putting the theoretical criteria discussed in the previous chapter into practice. Also, detecting entrainment is logically the first step of entrainment analysis - first one has to determine if entrainment has taken place, and after that, it can be analysed in more detail. In this chapter, I will discuss methods for detecting entrainment, using examples from three disciplines. Subsequent chapters are devoted to discussing methods for more fine-grained characterisation and analysis of both entrainment and interaction.

There have recently been three fields of inquiry where detecting entrainment has become important. These are a) un-intended coordination of body movements in interpersonal interaction, such as a conversation, b) comparative psychology, or the studies to investigate the scope of entrainment capabilities in non-human animals, and c) developmental psychology, especially the investigation of the development of entrainment abilities in young children and even babies. I will give examples of studies in these three fields,

focusing on the methods of detecting entrainment, thresholds etc.

These three fields are methodologically relevant for this thesis. Studies of unintentional coordination are examples of behavioural synchrony, discussed also in the previous chapter. These studies illustrate the wide range of behaviours that can automatically entrain or converge, also reminding that the extremely high coordination and synchrony in many musical behaviours is just a specific manifestation of a more general phenomenon. Comparative studies are important in investigating the evolutionary origins of musical behaviours, and human capacity for culture more generally. In order to trace the probable phylogeny of these capacities, assessing the capacities of non-human species, both our nearest relatives, other primates, as well as more distant species such as birds and sea lions has been undertaken. Synchrony detection -studies are required to analyse whether a certain animal has that particular capacity or not; a more detailed inspection is not often possible, nor necessary.

In addition to phylogeny, it is of interest to trace the ontogeny of these capacities. Understanding how the capacities for social interaction develop is important for being able to better understand developmental disorders such as autism, or disentangle the biological and cultural influences to these behaviours. Developmental psychology often investigates at what age a certain capacity or a behaviour emerges. To this end, the detection approach is a suitable tool. In the following sections, I will in turn review studies from each of these three fields, paying attention to how they measure entrainment and how they set the criteria for successful entrainment.

3.2 Unintentional coordination

Interpersonal coordination, or interactional synchrony (Bernieri, Davis, Rosenthal, & Knee, 1994) is one aspect of “rapport”, and as such it indexes various social psychological processes. For example, unintentional mimicry and synchrony of body movements has been suggested to support the idea that our understanding of the actions of others is based on simulating them in our mind. Interpersonal rhythms have an important role in the coordination of communication, and Adam Kendon (1970) suggests that it is not merely a vehicle for a smoother turn-taking but it serves a communicative function in itself. This is also evident in the early interaction of infants and caregivers

(Malloch & Trevarthen, 2009)¹.

Body mimicry and coordination have usually been studied by annotating videos and coding videoed behaviours, as in Bernieri, Reznick, and Rosenthal (1988). This is obviously a very time-consuming and limited method, as you can only code what you can see. I will demonstrate in a later chapter that using motion capture (see appendix B), coordination can be measured from movement that is not visible (or at least not reliably codable) by the naked eye. Also, as opposed to coding that is in practice limited to just a few features, by using motion capture body movements can be recorded with high dimensionality and high precision. The studies discussed below mainly use some form of either optical motion capture or other direct measurement of movement.

Interpersonal synchrony has been studied as an example of spontaneous pattern formation (Oullier, de Guzman, Jantzen, Lagarde, & Kelso, 2008), and in general the dynamics of these two-person systems has been studied since the 1980's. Early studies, like the one by Haken, Kelso, and Bunz (1985), involved abstract laboratory tasks like finger-wiggling or pendulum-swinging. More recently, these dynamics have also been examined in more ecological contexts. In these more naturalistic settings, participants can be studied while having natural conversations, walking and talking side by side, or sitting in rocking chairs (Kugler & Turvey, 1987; Nessler & Gilliland, 2009; M. Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; R. C. Schmidt & O'Brien, 1997). In these settings, body sway or gait entrainment serves as an index of interpersonal coordination or communication.

3.2.1 Pendulums, rocking chairs, and treadmills

A classic set-up for studying interpersonal coordination has been the Kugler & Turvey method (Kugler & Turvey, 1987), where participants are seated side-by-side, and they are swinging pendulums (e.g. R. C. Schmidt & O'Brien, 1997) and either attempt to synchronise or not. While this method yields good data and swinging a pendulum is easy to do for a sustained period of time, and it does not require constant monitoring or attention, it is not a very naturalistic task. While the pendulum is an external "instrument", it

¹Which ties this field together with the developmental research, and as it has been argued that these skills are innate, also the evolutionary theories come into play.

does not resemble a musical instrument and so the ecological validity of such a setup can be called into question. This point was also made by Richardson et al. (2007, 869), and they argued that a pendulum does not capture the “incidental movements that might be the basis for synchrony in naturalistic interactions”.

So, they suggested using rocking chairs instead, and in their own studies they investigated both intentional and unintentional coordination of movements (M. Richardson et al., 2007). In their setup, pairs of participants sit in rocking chairs that can be weighted so that their natural rocking frequencies can be manipulated. After demonstrating that side-by-side rocking in rocking chairs follows the same coordination dynamics as hand-pendulum swinging (Kelso, 1984) and other self-organised coordination, they proceeded to study unintentional coordination. They wanted to see if they could manipulate the strength of visual coupling and also modulate the coordination tendency by weighting the chairs.

Covering the real purpose of their study, they asked pairs of people to rock in their preferred tempo, and focus their eyes to either straight ahead (having peripheral vision of each other’s movements), away from each other (no vision) or towards the other participant’s chair (focal vision). According to their hypotheses based on earlier studies, they found that both the weights and the visual feedback condition had a significant effect on the interpersonal coordination. The more strongly the visual information was focused on the partner’s movement, the tighter the synchrony.

Richardson et al. were working with movement data that was collected using magnetic Polhemus trackers that were attached to the headrests of the chairs (Richardson et al., 2007, 872). After centering the data series to zero and low-pass filtering it, the authors proceeded to calculate the phase of the two chairs’ movements, from which they then determined the phase difference between the two chairs, the phase shifts and standard deviations of the phase differences.

In Richardson et al.’s study, using rocking chairs allowed global body movements be studied. Rocking, as opposed to postural sway, is also clearly periodic rather than stochastic, which makes synchronisation detection easier (Richardson et al., 2007, 869). As the movement data comes from a marker attached to the chair rather than the person, the movement of the marker is much more constrained and the data therefore needed less filtering.

Gait entrainment often occurs spontaneously and unintentionally. Two people walking side by side often lock in step. This form of unintentional interpersonal entrainment has also been studied in the lab with the help of treadmills. An experimental design where the effect of having a conversation was investigated was used by Guy Hayward in Cambridge. In Hayward's study, the data was manually converted to discrete MIDI format, so that it could be analysed as tapping data, and the analysis followed the course presented in the section about circular measures (5.1). Participants' steps were more entrained when they were having a conversation, but this was not a statistically significant result, probably due to the small number of participants.

Nessler and Gilliland (2009) tested the effects of leg length differential and altered sensory feedback on gait entrainment using two treadmills and pairs of people walking side by side. In their altered sensory feedback conditions, they restricted either the vision, sound, both, or neither, so that the two participants walking side by side had different strengths of sensory coupling. In this experiment, walking was the only task, there were no interaction or distraction tasks. In addition to the four visual & auditory conditions, they had an enhanced tactile condition, where the two participants were mechanically coupled together from their waists using a velcro-pvc-pipe-spring contraption. And finally, they were asked to intentionally walk in synchrony as closely as they could.

The authors correctly note that some degree of period and phase locking can occur simply by chance (Nessler & Gilliland, 2009, 776), and they suggested a smart way to take this into account in their analysis. To estimate the amount of *chance entrainment*, they also asked the participants to walk individually. Entrainment measurements of combinations of these individual trials, where obviously no entrainment could have taken place, was then added into the ANOVA model as a baseline. The other conditions were compared to this baseline to establish whether they exhibited a higher than chance level of entrainment.

The trials lasted one minute each and step data was collected using an optical motion capture system and ankle markers. Period and phase locking were analysed separately. Their criterion for period locking was that the step frequencies (cadences) had to be within 0.02 Hz from each other. Cadence was calculated in a 5-second moving window. From this data, they

also calculated the percentage of frequency locking per each trial, and the mean frequency difference between the two participants. Their criteria for minimum entrainment was that if at least 10% of time (5.5 seconds, about 4 steps) the pair exhibited frequency locking, then the pair was considered to have met the condition and was selected for phase analysis.

The phase analysis was performed in much the same way as is proposed in this thesis for discrete tapping data. The duration of each step (analysed from marker position data) was calculated and then phase was posited to grow linearly from 0 to 2π during each step. The continuous measurement allowed for continuous phase relationship measurement, instead of just a discrete one (once per step). Thus the number of data points is much higher, which helps in detecting possible differences between the sensory conditions, as these differences are likely to be very small.

The authors also had minimum criteria for phase locking. They considered both in-phase and anti-phase entrainment, and if the phase difference between the two participants was within 10 degrees (0.175 radians) of either 0 or π , they were considered to be phase-locked. Similarly to period locking, the metric for statistical analysis was the percentage of phase-locked movement of the total movement. Note that phase locking was only considered if there was period locking as well.

Their results match expectations very well. The leg length differential was found to have an effect on entrainment: participants whose legs were of similar length were more entrained. In the intentional entrainment trials, all pairs achieved frequency locking at some trials, and frequency locking was observed in 98.9% of these trials. The frequency locking rate was much lower in the “altered sensory conditions” (44.9 – 62.3%). The differences between these conditions did not yield statistical significance, but the results are interesting as trends. In terms of frequency locking, the mechanical coupling yielded the highest rate, followed by the unaltered condition (participants had unrestricted vision and sound). As in tapping studies, having unrestricted sound only was better than only being able to see each other. The condition with both vision and sound restricted was the lowest, as expected.

In terms of phase locking, the results get a bit less clear. This is partly due to the analysis workflow, where the data is first filtered by accepting only those trials where frequency-locking is observed, and then phase-locking is accepted only if it falls within a very narrow 10 degree window either in-phase

or anti-phase. Another option would have been to measure the amount of phase coherence directly, instead of setting these criteria and calculating how many trials would pass them.

However, looking at both in-phase and anti-phase entrainment, mechanical coupling yielded the highest phase locking rates, followed by the sound only -condition and normal sound and vision -conditions. Visual only and the condition with both modalities restricted were again worst. These results are similar to those obtained in my pilot studies. In general, sensorimotor entrainment through sound is tighter than through visual connection. (See appendix A for more details.)

In Nessler and Gilliland's study, the mechanical coupling turned out to be stronger than the auditory or visual ones. Having the mechanical coupling at the waist did however leave enough freedom for the participants, as was evidenced by the lower period and phase-locking rate compared to the intentional entrainment condition. An even tighter coupling, say, mechanical coupling of the legs, would probably have resulted in extremely high entrainment rates (or participants tumbling down), but that would not "count", as it would violate the condition of independence (see chapter 2.2).

The study by MacDougall and Moore (2005) provides a more general explanation for this phenomenon. In their study they showed that humans have a strong tendency towards a 2 Hz resonant frequency, so for example, that is the preferred cadence when walking. They also showed that this is not dependent on participants' gender, age, height, weight, or body mass index. This tendency would in part explain at least the period matching that Nessler & Gilliland observed, and at least the 8.4% of chance synchrony. As most people have a very similar gait cadence, it is easy to entrain with others as very little adjustment is needed. Thus most people will be within each other's entrainment region, which helps explain why gait entrainment is so general, and how it can be easily appropriated for military purposes. (McNeill, 1995).

3.2.2 Body sway and posture

Shockley, Santana & Fowler (2003) studied coordination of body posture in conversation. In their experiment, pairs of participants performed the task of comparing small differences in cartoon pictures with each other. The participants were standing, and were either conversing with each other and

facing each other (verbal + visual condition), conversing with each other but facing in opposite directions (verbal only condition), facing each other but conversing with a confederate of the experimenter and not each other (visual only condition), or finally, were conversing with the confederates and facing away from each other (no contact condition). While having these conversations, the participants had a movement sensor attached to their waist so that changes in their body posture could be measured, and the inter-individual coordination quantified.

The body movement Shockley et al. were interested in was body sway. This sway is largely sub-conscious, and is not only needed to maintain balance in upright posture, but also serves as a foundation for “suprapostural” activities such as pointing, grasping, gesturing and conversing (Shockley, Santana, & Fowler, 2003, 327). This movement poses a number of challenges for the analysis. It is often small in amplitude, which means visual observation is not sufficiently accurate method for studying it, but movement sensors are needed. The movement itself is irregular, non-periodic and non-stationary, meaning that the time-series is very complex and thus “traditional” synchronisation analysis methods, auto-correlation analyses etc. will not work very well.

Thus Shockley et al. used magnetic Polhemus trackers to capture the body movements of the participants. In their data analysis, they used cross-recurrence quantification (CRQ, see chapter 5.2 and appendix D). Using this method, the first step is the reconstruction of the phase space of the two processes (say, x and y). After that, the joint structure of those processes can be made visible by plotting a point on the x-y coordinates, when x_i and y_i fall within a certain pre-defined radius from each other. This cross-recurrence plot is then quantified for statistical analysis, and measures of overall synchronicity, stability of the relationship, length of longest stretch of synchrony etc. obtained.

In the analysis, these measures were then subjected to a traditional statistical analysis of comparison of means and variance between groups. The team found that the body movements of the two participants were coordinated when the two were in actual conversation (verbal conditions) regardless of the visual condition, i.e. whether the two participants could see each other or not. The coordination was much weaker when the participants could only see each other but were having a conversation with a confederate, rather

than each other. This suggests that information exchange is pivotal for this body sway entrainment, and it is not based on just the visual exposure and subconscious mimicry of the person that one sees. This is a very interesting result, suggesting that low-level behavioural synchrony helps maintain a readiness for higher-level joint action.

With their clever experimental design, Shockley et al. managed to avoid the need to set thresholds for entrainment. Rather, they clearly defined the metrics they were going to use to quantify the phase relationship of the two participants, and then statistically compared performance in those metrics across different conditions. This allowed them to conclude that verbal interaction had an effect on anterior-posterior body sway, a noisy, non-periodic, non-stationary signal.

It is important to note here that we would expect musical or quasi-musical movement patterns (dancing, tapping etc.) to exhibit much more structure (e.g. periodicity) and coordination than these un-intentional, sub-consciously occurring movements. Therefore, CRQ provides a useful set of tools for detecting small changes in coordination of close-to-chaotic signals, such as “interpersonal postural coordination”, but these are not the best choice when studying intentionally coordinated movements.

3.2.3 Summary

In studies that have looked at unintentional entrainment between people, it has been observed constantly that we do tend to entrain body movements with our walking and conversation partners. This interpersonal entrainment can be used as an index of social interaction, as it occurs between people who exchange information and communicate, rather than just based on visual exposure (Shockley et al., 2003). While it is fascinating that these subtle movements can be measured (either directly as Shockley et al. did, or using a proxy movement such as gait, pendulum or a rocking chair) and that synchronisation can be detected from the data, the real issue is, what does this tell about our cognition.

This kind of behavioural synchrony has a facilitating role in interpersonal interaction. Behavioural synchrony can be a residual effect of the mutual coupling of action-perception mechanisms, where action understanding is based on co-activation of the motor system of the observer, although direct imitation of the observed movements is suppressed (Prinz, 1997; Sebanz,

Knoblich, & Prinz, 2003). This subtle entrainment could serve the function of keeping the communication lines open, maintaining readiness for joint action for the dyad ready for joint action. In that sense, it might serve a “phatic” function in communication (Coupland, Coupland, & Robinson, 1992).

From the point of view of enactive, embodied research (e.g. K. L. Marsh et al., 2009; Vallacher & Jackson, 2009), it has been argued that these kinds of results demonstrate the fundamentally embodied nature of cognition. The idea is that complex social interaction emerges from simpler processes such as entrainment. The action-perception equivalence mentioned above can be seen as similar evidence of the more complex function of *action understanding* arising from these simpler resonances and entrainments. The cognitive explanation of, for example, Sebanz et al. acknowledges that such synchrony takes place, but in their cognitive models it subserves higher-level mechanisms, such as mental representations and simulation processes. The enactive view is that the emergence replaces constructs such as mental representations.

These differences between the approaches can at least partly be a result of focusing on different time scales. The embodied and enactive accounts often argue that the cognitivist models of social interaction (see for example Knoblich and Sebanz (2008) or Tomasello et al. (2005)) are too complicated and consequently too slow to account for the immediate and automatic mirroring and mutual adjustments that are observed in dyadic and group interactions. On the other hand, the cognitivist theorists complain that the enactive view and emergence are too simplistic and can not account for the richness and complex features of interaction, how for example our beliefs about our partner’s beliefs influence our behaviour. Both are good points, and it is possible that both theories just focus on different types of interactive behaviours, and they both can be useful. It is likely that we have a system for extremely fast social reactions, for automatic perception of social intentions and adjustment of behaviour. This system would be seen to operate when we entrain to the rhythms of our conversation partners, mirror their gestures, or adjust our posture when for example carrying a table with someone; the automatic, immediate, “on line” attunement. In contrast, many social interactions require a different system. This system powers the more sophisticated, conscious, and controlled “off-line” planning

and analysis of the partner’s intentions, what they are likely to know and not know; processes involving a theory of mind and abstract thinking that is not based on information available in the situation. Complex forms of social communication (such as language and music) might include both types of processes (see chapter 2).

This division of cognitive processing into two different systems is not a new idea, one such division in the field of decision-making has been discussed in great clarity and detail by the Nobel-prize winning psychologist Daniel Kahneman (2011). Kahneman proposes we have a fast, intuitive system and a slower, more rational system. Whether the fast, automatic social coupling and the slow, theory-of-mind -based social evaluation are linked or somehow mapped to Kahneman’s two systems is unknown, as there has not been systematic studies on the timescales of interaction.

3.3 Comparative studies on synchronisation and entrainment

When studying music as a biological as well as cultural phenomenon (see section 1.2), it is helpful to investigate the evolutionary origins of our musical capacities. Alongside genetic and fossil studies, comparative behavioural studies with non-human species are an important tool in this pursuit. For example, comparing human capacities to those of other primates, our closest relatives, can provide evidence about the phylogeny of those capacities. Both continuities and discontinuities are informative. Finding homologies for these capacities in our more distant relatives, across larger phylogenetic distance (for example, across different vocal learning species—see below, section 3.3.3) can be very informative about the neural, genetic, as well as functional nature of those capacities (I. Cross et al., 2013). Entrainment and sensorimotor synchronisation have been studied from this comparative perspective, and in the following sections I will look into the methodological aspects of these studies.

Is sensorimotor synchronisation to an external referent a uniquely human skill²? Recently, anecdotal evidence in the form of YouTube-videos has been cropping up, as well as studies investigating the entrainment abilities of

²It is important to remember the distinction made between sensorimotor synchronisation and interpersonal entrainment in chapter 2.1.2.

birds such as Snowball (Patel et al., 2009), elephants (Patel & Iversen, 2006), rhesus monkeys (Zarco, Merchant, Prado, & Mendez, 2009) or vocal learning species in general (Schachner, Brady, Pepperberg, & Hauser, 2009). These studies would benefit from clear, agreed upon criteria for detecting entrainment. Such criteria have been proposed along the lines of section 2.1 in this thesis. Unfortunately the actual research settings have different constraints and practical limitations especially given that behavioural animal research is very difficult so that the experimental designs are often far from perfect. And, in these studies it is usually not possible to circumvent the need for entrainment thresholds by clever experimental designs, either. Often researchers have just had to analyse pre-recorded footage. Occasionally, perturbations or manipulations of the referent have been used as part of quasi-experimental research designs, but to my knowledge, in no study have the conditions for interaction been changed, by manipulating the affordances for interaction of the setting (e.g. comparisons between mutual and one-sided communication).

The issue of animal entrainment has been investigated in detail by John Bispham (Bispham, 2006), and I do not intend to repeat his argumentation in full, or delve too deep into the vivid discussion surrounding the issue. Rather, my interest in this field is mainly methodological, and in the following sections, I aim to summarise and evaluate the methods and synchronisation criteria used so far. The mechanistic, collective synchrony in crickets, fireflies or frogs is different from human entrainment (Strogatz, 2003). Humans are flexible in terms of rhythmic complexity of the referent, and able to synchronise in complex $m : n$ patterns; humans can synchronise with stimuli across a relatively large range of tempi, and humans are able to synchronise motor or vocal outputs with sensory stimuli in many modalities (sensori-motor synchronisation). This differs from the simple, holistic oscillatory behaviour of fireflies, for example, which in terms of mechanisms is closer to the entrainment of clock pendulums. In the following sections I will discuss some examples of non-human entrainment and synchronisation that are (potentially) similar to the human capacity.

3.3.1 Primate studies

In a comparative study investigating the synchronisation and beat detection, Wilbert Zarco and Hugo Merchant with their collaborators studied rhesus

monkeys and humans (Zarco et al., 2009).

The monkeys were practicing tapping trials for up to 25 months (!), and in that time learned to react to auditory and visual stimuli, but not to engage in anticipatory action that is typical of human synchronisation. The monkeys were also somewhat more variable in their tapping, but this major difference, even in spite of extremely extensive training, shows that the human rhythmic behaviour is qualitatively different from what the monkeys could achieve.

In a follow-up study, Honing et al. (2012) showed using event-related potentials and a mismatch negativity paradigm (Näätänen et al., 2011) that rhesus monkeys are able to detect pitch deviants and gaps, but they did not seem to be able to extract the beat out of the rhythmic stimuli, unlike human infants (Honing, Ladinig, Haden, & Winkler, 2009). This was an elegant and ethically sound way to demonstrate the difference between humans and these monkeys, without having to subject these animals to countless hours (6 days per week, average 4 hours per day, for up to 25 months) of drills in a meaningless task, being water-deprived and working for fruit juice drops, as in Zarco et al. (2009).

3.3.2 Snowball

Perhaps the most famous example of animal entrainment is “Snowball”, the sulphur-crested cockatoo who likes to move rhythmically (bobbing its head and stepping up and down and from side to side on the back of a chair or on its perch) to music, especially to the song “Everybody” by the Backstreet Boys. Snowball’s abilities were examined in some detail by Aniruddh Patel and his colleagues (Patel et al., 2009).

To test if the bird’s behaviour had characteristics of human entrainment, researchers constructed versions of its favourite song that were in different tempi. Thus, they could create a perturbation to the referent, and see if the bird would be able to re-synchronise its movements or whether the observed synchrony was due to a lucky matching between the its natural movement frequency and the original tempo of the music. This would also put its tempo flexibility to the test.

To measure synchronisation, Snowball’s performances were videoed and then manually coded. As always with movement data, it is first necessary to decide which aspects of the movement are going to be analysed. The researchers chose vertical head bobs, as they represented the most reliable

rhythmic gesture. Incidentally, vertical head movement is a good indicator of rhythmic body movements also in humans (Eerola, Luck, & Toivainen, 2006; Toivainen et al., 2010). The coders noted down the frame numbers where Snowball’s up-and-down movement would reach its lowest position, essentially producing a time-series of tap onset times. Similarly, the beat onset times were harvested from the music examples. Then, the researchers calculated the phase of the bobbing in relation with the beats in the audio, and assessed the synchronisation using circular statistics.

As Snowball was not consistently synchronised with the music, but in short bouts, the researchers decided to first identify these bouts by using a windowed analysis. First, those trials where Snowball did not move enough were dropped. The bouts were then identified using a version of the Rayleigh test that checks both period and phase alignment³. Then, only the synchronised data was pooled together and tested for mean asynchrony, which was close to zero, indicating synchrony with the musical beat (3.9 degrees)⁴. Finally, a permutation test and Monte Carlo simulations were carried out with full trials, and this analysis suggested that it would be very unlikely to get as many synchronised headbobs just by chance.

The really interesting part of the analysis is the distribution of these synchronised bouts to different trials. It turns out that only the slowest trials yielded zero synchronised bouts, while there were more in trials where the music was sped up.

However, as Snowball seemed to synchronise only at such short bouts, some questions about Snowball’s synchronisation abilities remain unanswered. For example, we do not know if Snowball could adjust to a gradually changing tempo or respond to perturbations in the musical beat. These would be stronger evidence for “human-like” rhythmic entrainment. Even though the analysis in Patel et al. (2009) is carefully conducted, it illustrates that passing statistical tests can only answer some of the questions in synchronisation analysis. As the study does not address Snowball’s error correction mechanisms, it remains an open question whether Snowball possesses a mechanism of rhythmic entrainment, or whether the synchrony is based on a more gen-

³There are some methodological concerns regarding using Rayleigh tests in short windows, I will discuss these in chapter 5.1.

⁴As only a subset of the data, preselected to have a small phase error was used, this is not surprising, rather, one could say it is a case of circular reasoning using circular statistics.

eral mechanism of interval-based timing coupled with energistic matching of movement to stimulus (see Merchant & Honing, 2014).

3.3.3 Vocal mimickry

Some species of animals, most famously songbirds, are capable of mimicking (learning) the sounds or “songs” of either other conspecifics, other species or the environment. In other words, these species are able to learn to produce complex vocalisations by imitating others (Kelley, Coe, Madden, & Healy, 2008). It has been suggested that this capacity is linked to the ability for “auditory-motor entrainment” (Schachner et al., 2009), in that vocal mimickry is a “necessary but not sufficient” requirement for entrainment.

Adena Schachner and her colleagues studied a gray African parrot called Alex and the cockatoo Snowball in detail, and a number of other animals whose videos had been uploaded to YouTube (Schachner et al., 2009). Their methodology was slightly different from that of Patel et al., in that they did not only compare the animals’ movements to beats in music directly, but also to human tapping on the same stimuli. This gave them a human baseline to which they could compare the animal performance⁵. In terms of synchronisation detection methods, both versions of the Rayleigh test⁶ (sensitive and insensitive to mean direction) were used. The authors also checked whether the animals’ overall movement displayed periodicities similar to those of the human participants. To this end, they calculated the number of pixels that would change from one frame to another in the video, which is a good index of movement. Then, via Fourier transform they broke this signal down to its frequency components, and finally used a Monte Carlo simulation to determine whether the peak they found was likely to be formed just by chance.

Schachner’s methodology is compatible with the criteria set in the previous chapter. They first of all check that there is periodic movement (Fourier analysis + Monte Carlo simulation). The movements of an animal are clearly *autonomous* from the production of auditory events of a CD player, or the production of movements by human participants, so that there is no danger

⁵They used the term “intentional entrainment” for what I would call “intentional synchronisation to musical stimulus”.

⁶Rayleigh test is explained in chapter 5.1.7. It is a synchronisation measure used with circular data; passing the test indicates that the two processes are synchronised.

of the two processes being either too strongly coupled or products of the same underlying process. The criterion for *interaction* or *coupling* is satisfied in that the animals and humans *hear* the music. The central tenet of the vocal learning -hypothesis is that the species who have this property have brain circuits that do have internal coupling between their auditory sensory systems and the motor output circuits. Thus, observing synchronisation would support this hypothesis. Finally, the authors set out to test the criterion of *flexibility* by presenting the stimuli in various tempi, to rule out coincidental synchrony and to map out the potential entrainment/synchronisation region. Unfortunately they were able to use this procedure for only one of the two actual experimental subjects, and of course this manipulation was not possible for the performances in YouTube videos.

In the end, detection of synchronisation came down to two measures: a) maintaining constant phase and b) matching in frequency, and understandably so, as the aim of the research was to find *any* evidence for entrainment. Of their two actual subjects, Alex the parrot managed to maintain a constant period in all trials, but failed to maintain consistent phase relationship in most and failed in matching phase in all but one of the trials. Snowball was more successful, managing to meet the criteria in different tempi and with head and feet movements. The metrics that the authors present are a proportion of the human participants that match the phase of the birds with $p = 0.05$ significance level.

The YouTube video data was combed through with the same methods. Here the criteria were relaxed so that meeting only one of them (phase or frequency) was enough. The main finding here was that all potential cases of entrainment were in vocal learning species, while none were found the videos of non-vocal learning species. This data suggests that some animals might indeed be capable of sensori-motor synchronisation, but offers no indisputable evidence. Because all of the potential entrainers were vocal learning species, this study also lends support to the theory that vocal mimicry is a requirement for such synchronisation. Also the primate studies discussed above in section 3.3.1 supports this theory, as rhesus monkeys, shown not to be able to synchronise with an auditory stimulus, are not a vocal learning species.

The final example comes again from Patel and Iverson, and their analysis of entrainment in a Thai elephant orchestra (Patel & Iversen, 2006). They demonstrated that the elephants (another vocal learning species) were

capable of producing a steady tempo, although it is not clear if this is mainly due to body ballistics (they were swinging their trunks) and whether there would be tempo flexibility. In this study as well, the evidence for entrainment is not strong, although bouts of synchrony were observed. Whether these bouts were due to chance or intentional was not systematically tested.

3.3.4 Summary: comparative studies

With all these results, the evidence for animal entrainment remains sketchy. However, in keeping with the classic black swan -example, just one demonstration of animal entrainment would be needed to sink the strong version of the theory that beat-based synchrony/sensorimotor synchronisation/interpersonal entrainment is uniquely human. But rather than boiling this issue down to a simple yes—no question it is more enlightening to look at details of this evidence and the meaning and function of entrainment in both animals and humans.

According to the current evidence, tempo flexibility in animal synchronisation is very limited. Also, synchronisation tends to take place in short bouts rather than in a sustained fashion, and to a limited set of stimuli. It is also not yet established whether only special individuals possess the ability to synchronise, or whether these capabilities are more general in the populations. The best evidence currently comes from animals who have lived their lives with humans and often in enriched environments.

As with other aspects of animal cognition, the inescapable limitation of the studies is that unlike with human participants, we have no common language and therefore no clear means to communicate our wishes and instructions as researchers. The field is a great example of the importance of remembering that absence of evidence can not be taken as evidence of absence. If an animal does not display certain behaviours, is it because it can not, or because it does not want to, or feel the need to do it? Treats and food are often offered as motivators, but such extrinsic motivators might not be enough, and we have a history of being too hasty to state things about animal cognition only to retract later when new observations prove otherwise.

As John Bispham (2006) has pointed out, the biggest difference between human musicking and any animal behaviour that even distantly reminding it is the motivational aspect mentioned above. Humans derive great pleasure

from music, rhythmic entrainment, and many are obsessed with it (Levitin, 2007). Our motivation to engage with each other through music and shared rhythms is universal. Musicking and entrainment also have various functions in human societies, and can even have had effects on survival (Huron, 2001). We are a social species with highly evolved needs for interpersonal coordination. What related functions could entrainment serve in non-human animals? Animal behaviours that to us sound like music (birdsong especially), or resemble human musicking, are usually linked to alarm, territorial or mating calls and similar situations. While we might be able to interpret them as music, the meaning for the animals themselves are different. The examples discussed above do not speak convincingly for intrinsic musical motivation, either.

3.4 Developmental studies

Tracking the developmental trajectory of a cognitive ability gives important information about that ability, and also helps disassemble complex behaviours into their constituent parts. The developmental trajectory of interpersonal entrainment also helps in investigating the role of entrainment in human sociality, in communication, music and language.

Devin McAuley and his colleagues studied synchronisation abilities across lifespan. They were more interested in discovering connections between age and various speed limits in tapping, rather than synchronisation as such. Professor Mari Riess Jones's dynamic attending theories pertain to the entrainment of our perceptual and attentional processes with the temporal regularities in our environment (Jones & Boltz, 1989; Large & Jones, 1999). According to her theory, regular rhythms in our sensory environment are important, and our attention is entrained to those regular pulses and rhythms, as an emergent function of the dynamic system that is our brain. According to the theory, the pulses that we are most likely to synch with are age-dependent. This study was then conducted to test competing predictions about the nature of this age-dependence.

In their study, the youngest participants were just four years old, while the oldest one was 95. They sampled the lower end of the age range (4–12 year olds) with bins of two years, so that they had approximately 20 participants in each. The majority of participants were 18–38 (119) and

further 52 were 39–59. There were also 25 60–74 year olds and finally 21 who were over 75 years old. Tapping tasks were set up to find the spontaneous motor tempo, preferred perceptual tempo and the fastest and slowest motor tempi. Synchronisation—continuation task was also performed to look at synchronisation abilities in different tempi, the inter-onset intervals of the metronome ranging from 150 to 1709 milliseconds.

Here, the performance in the synchronisation task was quantified in circular domain, using the mean angle θ as a measure of the average synchronisation error and the mean resultant length of the sum vector \bar{R} as the measure of consistency⁷. The comparison across tempi and age groups formed the main point here, and thus no thresholds for successful synchrony were used.

In their study, they found that the 150ms interval was too fast for anyone to synchronise with. The next tempo tested, 225 ms IOI was an interesting one. From the adult cohorts, the 18–38 year olds managed best, with performance declining with each older cohort. Of the children, the 10–12 year olds did approximately as well as the 39–59 year olds. Mirroring the adults, a linear decline was observed in the younger children.

Joëlle Provasi and Anne Bobin-Bègue (Provasi & Bobin-Bègue, 2003) tested children of 2.5–4 years old in tapping tasks, mapping their spontaneous motor tempo and also tested synchronisation to metronomes of different tempi. Their data indicates that already 2.5 year olds have a somewhat stable spontaneous motor rate, and they synchronise to a visually reinforced auditory metronome (they heard animal calls and had to press a plate, and if they managed to press within a predefined response window, a picture of the animal appeared on the screen). Synchronisation is not yet stable in these age groups and thus the typical tap-to-metronome synchronisation measures are not used, but rather inter-response intervals (tapping tempo) and IRI variation coefficients are used. The results in this paper suggest that children as young as this can produce a somewhat steady tempo, which then can be detuned towards an attractor (external stimulus).

Sebastian Kirschner and Michael Tomasello (2009) studied children in three age groups (2.5, 3.5 and 4.5 years), drumming with a human partner, a drumming machine and a playback drum track. They found that in all age

⁷These measures are discussed in detail in chapter 5.1. Similarity of these measures to the ones proposed in this thesis is not a coincidence. Devin McAuley initially introduced me to circular statistics and was very helpful in teaching me and the rest of the *Entrainment Network* how these measures work.

groups, children were more synchronised with the human partners than with the machine or tape. Even the youngest children adapted their drumming tempi in the social condition.

Methodologically, the problems with testing children's abilities are in many ways similar to those in animal research. Conveying instructions and feedback, managing attention and motivation and making correct interpretations of observations are all major difficulties, especially when working with infants, but also with the 4–5 year olds as in this study. A major issue is that finger tapping provides an abstracted response to the rhythm, and not a direct one. In other words, the child would not only need to be able to perceive the beat, but also be able to understand correctly instructions for a task that would not make too much sense, master fine motorics to be able to carry out the task and finally harness enough concentration to keep on going all through the trials. A lot of developmental research has been inspired by Piaget's theory of developmental stages (Piaget, 1952), even though the idea of set stages and especially age thresholds have been called into question, as more ecologically valid research paradigms have emerged (e.g. Bruner, 1964). Also, our understanding of the development of cognitive abilities related to music-making or rhythm perception have shifted and will shift through new, innovative research paradigms. While for adults, tapping tasks are trivially easy, and successfully completing these trials does not require musical training (although it helps), they might be a wrong method to study young children.

The early musical abilities of babies are often tested using various perceptual paradigms, such as measuring preferences with gaze direction paradigms (Phillips-Silver & Trainor, 2005) or brain imaging methods (e.g. Honing et al., 2009; Huotilainen, 2010). Using MEG or EEG bypasses the behavioural route altogether, allowing studying of even neonates, while of course these methods have their own set of limitations.

A smart combination of corporeal feeling of the beat and these observation methods was employed in the study by Jessica Phillips-Silver and Laurel Trainor (Phillips-Silver & Trainor, 2005). In their study, babies were bounced according to two different metric interpretations (duple or triple metre) of an ambiguous six-beat rhythm pattern, and then their preference of the two interpretations was tested using the headturn preference measurement paradigm (Trainor & Heinmiller, 1998), where the infant's preference

for a certain stimulus can be measured by the time s/he spends looking at the source of that stimulus. The results of this study show that how we move (or are moved) has in effect on how we hear things. This is a good example of embodiment of cognition in general and rhythm and music in particular.

A step forward in terms of embodiment were the studies by Tuomas Eerola and colleagues, where babies and toddlers were spontaneously moving to music, and these movements were then analysed for synchronisation (Eerola, Luck, & Toiviainen, 2006; Zentner & Eerola, 2010). Eerola, Luck and Toiviainen (2006) studied rhythmic engagement capabilities of 2—4 year olds using actual music and motion capture technology (see chapter 8 and appendix B). Interestingly, they found that any periodicities of the body movements were manifested in the vertical movement of a marker that was attached to a headband of the participant. The synchronisation detection analysis was based on windowed, enhanced auto-correlation. Autocorrelation will detect periodicities in the data (see chapter 6.4 for more details). The enhanced autocorrelation removes the “harmonics” of the periodicities that the algorithm finds, thus leaving just the “fundamental” frequency at which the movements are periodic. Calculating this in a moving window reveals the temporal evolution of the periodicity. Plotting this windowed, enhanced autocorrelation then allows you to see which participants exhibit which periodicities, and at which points of time.

Sharing the same period with the music is of course not yet synchronisation, as common period says nothing of the phase relationship of the movements and music, for instance. Eerola, Toiviainen and Luck say that in this particular study, phase locking occurred too rarely for them to be able to use that as a measure in the analysis. Instead, they looked at multiple tempi in an attempt to test the theory by Drake, Jones, and Baruch (2000) in that the range of synchronisation gets wider with age, and the preferred tempo gets slower. Their data follows this pattern, but the differences were not statistically significant. They did demonstrate, however, a methodology that can be used to study synchronisation abilities of children too young to tap.

Zentner & Eerola (2010) pushed this boundary even further. In their study, the youngest participants were just five months old (they had three age groups: 5—7 months, 9—12 months and 13—16 months). These babies were sitting in their caregivers’ laps and their movements to music and various

experimental stimuli, including speech, were analysed from either manually coded video data or motion capture data. Here, synchronisation measures were more indirect, as the authors measured whether different kinds of music had effects on the pulse clarity, average acceleration or velocity of the babies' movements. So rather than even attempting to see if the movements match to musical beats one-to-one, these more general links were investigated.

In the first part of their analysis, they showed that infants exhibited rhythmic movements much more when they heard music or other rhythmic stimuli than when they heard speech. Interestingly, infants found the version of Saint-Saëns piece where the rhythm was continuously rapidly fluctuating, so that one could not track the tempo, equally engaging as the original. To take a closer look at the rhythmic engagement, Zentner and Eerola calculated pulse clarity measures as well as kinematic measures (velocity and acceleration) of the infants' movements and compared those scores across different stimulus conditions. They found that infants do exhibit some degree of tempo flexibility, as faster tempi were associated with faster movements. This flexibility was shown to be clearest with metronomic rhythmic stimuli, and there were no differences between age groups, not even when the group was just split in two rather than three, using a median split.

3.5 Beat deafness?

The musical abilities that are discussed in this thesis are mostly such that are possessed by “everyone”, meaning that they are considered universal and do not require specialist musical training. These are the perceptual and cognitive abilities that allow people to engage with music and enjoy it. There are some people, however, whose basic musical abilities are impaired. According to estimates, about 4% of adults have amusia, a disorder mainly in pitch discrimination, and colloquially known as “tone deafness” (Henry & McAuley, 2010; Peretz et al., 2002; Stewart, 2006). The canonical tool for diagnosing amusia, the Montreal Battery for Evaluating Amusia, MBEA (Peretz, Champod, & Hyde, 2003) contains both pitch and rhythm-related sections. However, having difficulty in rhythm perception only is much rarer than having difficulties in pitch discrimination. Even though screening has been done online, only a few possible cases of “beat-deafness” have been found.

One such case is “Mathieu”, who was found in a recent screening of volunteers in Canada (Phillips-Silver et al., 2011). In studying his rhythmic abilities in more detail, in addition to the MBEA and the WAIS intelligence tests, the researchers performed other tests for rhythm perception and production. The production tasks were dancing (bouncing) to the beat of either a metronome or merengue music, or with the experimenter without music. The analysis was based on acceleration data recorded with a Nintendo Wii remote controller strapped to the subject (see appendix B).

The analysis of Mathieu’s dance movements looked at both period-matching and phase-locking. Vertical acceleration data was filtered and its instantaneous phase was calculated using a Hilbert transform. Then, this instantaneous phase was sampled at the times of the music beat onset, to see what the phase of Mathieu’s movements were at onset times of the beat in the music. In other words, the researchers performed stroboscopic observation of the movement data, where the strobing was determined by the phase of the beat sequence in the music (Pikovsky 2001, 62 & 163–166). The obtained phase values were then analysed using circular statistics and the Rayleigh test was calculated to determine whether the distribution of phase angles (between Mathieu’s movements and the music’s beat onsets) differed from uniform. If Mathieu’s movements are synchronised with the beat, then the instantaneous phase of his movement will always be similar at times of the beats, the distribution unimodal and narrow, and the Rayleigh test score high. It turns out that for some pieces of music, the ‘techno dance’, ‘pop dance’ and ‘dance rock’ pieces the Rayleigh test gives a positive result, indicating synchrony between movements and music. For most pieces, however, the test score is low and the result not statistically significant. Remembering that passing the Rayleigh test is rather easy, the results are in line with the hypothesis that Mathieu has unusually large problems synchronising movements with music. As he seems to be OK with the metronome and with musics with a very salient, emphasised beat (such as techno dance), this points out to a problem of abstracting the regular beat from within the musical sound and structure. Further studies with Mathieu will hopefully reveal the source of these problems or at least clarify the specific nature of them.

This analysis shows a good example of how the synchronisation of a continuous signal (movement data) with a discrete signal (beat onset times in music) can be analysed. Tools such as the MIR Toolbox (Lartillot &

Toiviainen, 2007; Lartillot, Toiviainen, & Eerola, 2008) can be used to automatically extract beat onsets from an audio file. After that, stroboscopic observation of the instantaneous phase of the continuous data can be carried out, and circular statistics used to analyse entrainment.

3.6 Summary: detecting synchrony and entrainment

While we are used to measuring the strength and stability of synchronisation in rhythm and timing research, there are many cases where it is instead necessary to simply detect whether there is entrainment at all. For example, when looking for evidence for entrainment abilities in either non-human animals or very young babies, this is a good approach. Or, this detection strategy might be useful at the edges of the spectrum of human timing abilities, as the case of “Mathieu” demonstrates. This could also be a useful approach for analysing music therapy interactions. Applying these sensitive detection methods on body movement analysis outside musical contexts has helped uncover that people synchronise unintentionally when they communicate with each other.

Both comparative and developmental studies suggest that in animals as well as very young children, interaction is based on “energistic” aspects rather than on beat-based entrainment. Rather than predicting the timing of the next beat, the response is based on matching the energy (arousal) and valence of one’s movements to those of the stimulus. This leads to a crude tempo matching but not necessarily a constant phase relationship. If the tempo range of the movements of the animal or the child is small, and coincides with the tempo of the music, this leads to at least bouts of synchrony. Taking this “chance synchrony” into account when designing detection studies is crucial, especially if one wants to make the “stronger” claim of sensorimotor synchronisation and not just the “weaker” claim of energistic interaction. A tell-tale sign of synchronisation is the presence of error correction that helps maintain synchrony over longer periods of time or re-adjust the sync when it has deteriorated.

Methodologically, there are many options available for these studies. In some studies, a threshold has been set and then performance measured against that. An example of this was Patel et al. study with “Snowball”

(2009), where Rayleigh test (see section 5.1.7) was used to determine whether synchronisation was statistically significant. In most other studies discussed above, no absolute threshold was set, but the presence of synchronisation was measured in a relative way. For example, in Shockley’s body sway study (2003), body sway synchrony was compared between a pair that was engaged in a conversation and a pair that merely was facing each other but not conversing together. In the case of “Mathieu” (Phillips-Silver et al., 2011), his performance was compared to a control group; in Zentner & Eerola’s study on infants (2010), the comparison was made between infants moving to speech and moving to different musics. These relative designs avoid the problem of having to set an absolute, predetermined threshold. Of course, it is not easy to imagine what a good control would have been for Snowball’s performance, for example.

All synchronisation detection methods are based on finding correspondences either in the period or phase of the two processes that are investigated. Obviously, linear methods that can be utilised in analysis of synchronisation-continuation tapping trials are not usable when such one-to-one correspondence between event onsets can not be assumed. There are different methods for continuous and discrete data, and the two data types can also be combined, as in the study by Eerola, Luck, and Toiviainen (2006). Their synchronisation can be measured using stroboscopic observation of phase of the continuous process, at onset times of the discrete process (Pikovsky et al., 2001). Cross-recurrence analysis is somewhat difficult to use, but will work even when synchronisation is very weak and the data noisy and non-stationary. Circular statistics can be used either for obtaining indexes of entrainment or stability (as in chapter 5, or in Zentner & Eerola, 2010) or to make inferences of the distributions of phase angles obtained (as in Patel et al., 2009).

These examples of studies where the objective is to determine whether entrainment occurs or not, are good reminders of the wide variety of contexts for entrainment. The studies in unintentional behavioural synchrony illustrate that this automatic and effortless mechanism is linked to all kinds of communicative behaviours and contexts. Comparative and developmental studies reflect the efforts to understand the phylogeny and ontogeny of rhythmic entrainment. Within these fields, debate over proper definitions and criteria for entrainment continues, as well as discussion over thresholds,

analysis methods and interpretation of experimental results. To exaggerate somewhat, changes in any of these can change the result of a study from “yes” to “no”, so this debate is a crucial part of advancement of research in these fields. There is currently some consolidation in methods and metrics in the field, but not yet a consensus.

In the following chapter I shift the focus from detection of entrainment to characterising it or measuring the “amount” of it. I will review studies of coordination and entrainment in dyads, as well as larger groups, such as bands, ensembles and choirs. I will focus mainly on those studies that have looked at actual, musical contexts.

Chapter 4

Dyadic entrainment, ensemble timing, and leadership roles

4.1 Introduction

The vast majority of synchronisation studies are conducted on individuals and are based on linear error correction models developed by Wing and Kristofferson (1973), and refined later by e.g. Schulze and Vorberg (2002). Bruno Repp, having authored a significant number of these studies, has also written two excellent reviews of synchronisation literature, first in 2005, and another in 2013 for studies done after the first review (Repp, 2005a; Repp & Su, 2013). These studies have examined many aspects of sensorimotor synchronisation, mainly using finger tapping as a response. This research has mainly looked at error correction in synchronising with a metronome, and time-keeping in continuation tapping after the metronome has been faded out. This body of research has been so well covered in Repp’s reviews that I will not revisit it in more detail here. Instead, in this chapter I will look at the few studies where there have been two participants entraining with each other, or where the pacing stimulus has not been a rigid, pre-programmed metronome. In accordance with the terminology adopted for this thesis, these are experiments where there can be entrainment, rather than just one-sided synchronisation.

I will start with dyadic tapping studies, similar to my own cooperative tapping experiment reported in chapter 7. After that, studies of ensemble timing are discussed, followed by studies of orchestra conductors. In musical

ensembles, bands and orchestras, participants have various roles. It will be shown that the stereotypical roles of “leader” and “followers” are not so straight-forward in musical interaction, at least when looking at timing data. After discussing these roles, I will present studies that look at entrainment in larger groups.

4.2 Entrainment in dyadic tapping studies

Modelling is a good way to investigate the properties of theories and also to test their validity. The linear error correction model has been put to such a test by Repp and Keller (2008). Their method was to systematically vary one parameter of the linear error correction equation that is supposed to keep the computer tapper synchronised. This parameter is α , a phase error correction coefficient that determines the proportion of perceived asynchrony that is corrected by the next beat. In other words, it represents the “gain” in the error correction function. Negative values of α lead to adjusting the beat timing to the wrong direction, increasing the asynchrony; zero α leads to no correction, α of 1 means that all of the error is corrected immediately, and values larger than 1 lead to overcorrection. In other words, α characterises an aspect of phase error correction, how efficiently the tapper (or the computer) reacts to errors. It is assumed to remain constant and it is not something a person for example would consciously control, as phase error correction is an automatic process.

In Repp’s and Keller’s experiments, human participants tapped with a computer tapper, and α was systematically varied between conditions. They even tried an “uncooperative” setting where the computer would adjust their timing to increase the asynchrony (negative α values). The main finding was that musically trained participants managed to maintain synchrony with the computers over a wide range of parameters. This illustrates the flexibility of human sensorimotor synchronisation. Repp and Keller also analysed the tapping data to determine the parameters of the human participants’ error correction, especially what their α was. They concluded that participants had a constant α throughout the trials, and varied the computer tapper’s α so that the joint α was approximately 0.9, a value that corrects almost all of the occurred asynchrony but avoids overcorrection.

However, as the computer α was kept constant throughout each individ-

ual trial in this study, we can not conclude that humans would maintain a constant phase correction gain when in actual interaction. Even in this study, participants varied their α in response to what the computer setting was, and so in dynamic interaction between two participants that are free to vary their gain, participants might end up changing it on the fly. Most methods Repp and Keller suggest for estimating α from tapping data assume that it is a constant that is calculated based on autocorrelation over a longer segment, or other averaging measure. I conducted one attempt to estimate α values from dyadic tapping data, but all suggested methods produced conflicting and unreliable answers (Himberg, 2011). This could be due to added dynamism and complexity that having two adaptive agents in the interaction brings. They might not just correct errors linearly like in the model, but tune the parameters of their error-correcting apparatus, as well. Nonlinear models are probably better suited for modelling these systems, as they provide the flexibility and dynamism that is required (Large, 2000).

It could as well be that a model that has virtually no memory of its past (the timing of a “tap” is adjusted with information only from the preceding tap) might not be an adequate one, at least it differs a lot from what the human *experience* of synchronisation is. Our easy, anticipatory and flexible synchronisation is experientially effortless and often described as “feeling the beat”. This feeling is lost when tempo is slowed down enough (ITI of about 2 s.). As a consequence, behaviour shifts from anticipatory to reactive. Similarly, when tempo is sped up or slowed down, the memory of interval lengths spans shorter or longer sections, as it does in expressive timing, where timing patterns span eight bars and even larger structures of music (E. Clarke, 1982).

Of the few dyadic tapping studies, Konvalinka et al. (2011) uses methods that are essentially identical to those presented in this thesis. In the study, the main observation was that the inter-tap intervals of the participants were not correlated at a zero lag (as would be expected), but they were instead correlated at lags 1 and -1 simultaneously. This is also surprising, as one or the other would be a sign of that participant leading (and the other following, correcting their timing one beat behind). But both to occur at the same time was not as expected at the time. This result has cropped up in a number of studies since, including a number of my own pilots and experiments, and the interpretation is that it is a sign of mutual adaptation—both participants

constantly adjusting their performance in response to what they do, what the other person is doing, and what the emerging joint performance is. Mutual adaptation has conceptually been discussed in communication studies and interaction studies for quite a long time (see e.g. Cappella, 1991), but had not been measured quantitatively in such a systematic fashion until very recently.

The mutual adaptation hypothesis was also investigated by Nowicki, Prinz, Grosjean, Repp, and Keller (2013). In their study, participants were tapping joint and solo trials, and in the joint trials they were tapping in syncopation with each other. They measured mutual adaptation by looking at joint lag 1 autocorrelations of asynchronies between the participants. In their experiment 1, the participants were synchronising with a pacing signal and syncopating with their partner, making it a mixed task. The partner's tapping did influence the other's timings, however, in what the authors called "interpersonal temporal assimilation". This means that while the other tapper's timing errors influence the taps, they do not disrupt the synchronisation with pacing signal completely, rather, the earliness or lateness of the partner is assimilated into the timings. This assimilation also means that the timings are shifted to the opposite direction than if they were "compensating" for the partner's error, as the linear model used in Repp's and Keller's study does.

So far all studies of actual interaction where two agents can adapt to each other and proper entrainment can occur, have produced "surprising" outcomes that do not seem to fit the linear models typical of synchronisation studies. Often, interactions that emerge are more noisy and less predictable than usually observed (or tolerated) in the synchronisation studies. Mutuality and bidirectional effects are commonplace. I will return to these studies when discussing their methods in more detail below, in chapter 6.

4.3 Ensemble and group timing

Rudolf Rasch studied ensemble timing in a trio of trios (a recorder trio, a reed trio and a string trio) (1979, 1988). He made two recordings of each of the trios in an anechoic chamber using directional microphones, so that he could then analyse each of the instruments separately for onset synchrony. From these recordings, using an amplitude envelope graph to measure onset times of each instruments, Rasch analysed how synchronised the ensembles were,

calculating the “asynchronization of a pair of voices” for all combinations of voices. He learned that in the trios the main melody instruments tend to lead relative to the other instruments, followed by the lowest voices, and the middle voices trailing behind the others. He also found that the string trio had the largest asynchronicities (37 and 49 ms for the two pieces of music), while the wind trios were synchronised slightly more closely (around 30 ms). This is probably due to different rise times of different instrument sounds.

Rasch’s synchronisation measure, or the “asynchronization”, is root-mean-square of standard deviation of the relative onset times and onset time differences. The root-mean-squaring simply gets rid of the sign of the asynchronies, so that this measure does not distinguish between being ahead or behind, only the amount by which the onsets differ. Using the standard deviation of asynchronies rather than the relative onset time differences themselves, puts the focus on the stability of the relationship, thus taking into account that having absolute simultaneity of onsets is not always preferable; for example, the leading voice was found to be systematically ahead of the other voices, which is primarily a matter of perceptual salience, as this slight advance in time makes the musically most important voice also perceptually slightly more salient than the accompanying voices. This tendency for the melody to lead was also found by (Palmer, 1989), showing that when pianists play chord sequences, the melody note leads compared to the others.

Shaffer (1984) studied pianists and the temporal coordination of their left and right hands, as well as interpersonal coordination in a duet. He found that even with relatively little practise, pianists were able to produce consistent performances in terms of their expressive timing (departures from the notated timings), and very high synchronisation both within each player’s hands as well as between players. Shaffer used Rasch’s measure for ensemble synchronisation, and found that the asynchronicity in piano duos is similar to the asynchronicities Rasch found in string and wind trios.

Early research on ensembles was focused on the challenges of extracting onset times and developing synchronisation measures, and psychologically the process of interest was the internal clock, the timing apparatus of the mind. More recently, research on music ensembles has linked with the action-perception research discussed in chapter 2. For example, Keller, Repp and Knoblich (2007) also studied piano duets, like Shaffer earlier, but their research question was more specifically about what role does simulation play in

entrainment. They found evidence supporting simulation theory, as pianists synced more tightly with the playback they themselves produced than with the playbacks of others. In their study, pianists rehearsed both parts of a piano duet, then recorded one part and on a later occasion performed the other part with a recording of the first part. Keller et al. mixed the recordings so that they could compare pianists duetting with themselves and with others. They also asked their participants whether they recognised the recordings they made themselves. The hypothesis was that if we understand actions (and the consequences of actions, such as the sound patterns of a piano) by simulating the actions in our minds, the duets with ourselves should be coordinated the best, as our simulation of our own actions should be more accurate than simulating the actions of others. And indeed, the results indicate that pianists duet best with themselves, and are able to recognise their own recordings. These two also were found to go hand in hand, with a positive correlation between self-recognition and synchronisation.

A dyad is a special group; while it exhibits some features of groups, it lacks many of the more complex dynamics (Williams, 2010). As Rasch demonstrated (1979), having a third person in the ensemble makes things more complicated. But while studying dyads does allow studying actual interaction and the interpersonal processes, studying larger groups brings about new challenges. For example, there might be different subgroups of the larger group that entrain more tightly with each other than the group as a whole, or even competing subgroups all performing together, allowing within-group and between-groups entrainment to be compared. Much of music and dance are of course performed in groups larger than two, and many times the ensembles consist of several subgroups (such as different voices in the choir, different instruments or groups of instruments in an orchestra etc.). Eventually it will be necessary to continue on Rasch's path and expand the dyadic methods to be able to analyse entrainment in these larger ensembles.

In a dance hall or a disco, people move to and synchronise with a shared stimulus, while also to some degree synchronising their movements in larger groups. While it can be argued that this situation is mainly about everyone synchronising individually to the external referent, the music, there are also social cues from the other dancers and it can be argued that there is entrainment between them. These aspects were investigated by De Bruyn et al., who studied children and adolescents moving to various pieces of

music (De Bruyn, Leman, Moelants, & Demey, 2009). In their study, participants were dancing as small groups, either being blindfolded or not, thus forming the individual and social conditions. In this experiment, the authors found that in the social condition, the synchronicity of movements to music, as well as the movement intensity (mean jerkiness) were higher. In the second experiment, they wanted to check if the results were due to the blindfold and created the individual condition by using partitioning walls instead. This resulted in the disappearance of the synchronisation difference, but the mean intensity difference remained the same. Thus the social situation influenced music-related movements, mainly their mean intensity. The results were not as clear in terms of synchronisation to music vs. entrainment to other dancers, but they suggest that participants were “socially synchronising” to the music. Mutual entrainment, at the timing level or at a higher, gestural level, was not investigated in detail.

Groups have more complex dynamics than dyads. These were investigated by Maduelli and Wing (2007), using the example of a flamenco performance. They presented a network model of performers, informed by prior studies and their experience and understanding of flamenco. This theoretical model was then tested in the lab, by recording multiple performances in different combinations of performers (dancer, singer, guitarist and palmeros (clappers)). They analysed the performance cues used to regulate the performance, logging the frequencies of their occurrence. These cues were partly choreographed and partly ad hoc; partly for control and partly for expressive purposes; and partly they were used to maintain a status quo and partly to instigate changes in the performance. In general, the observed cues fit very well their network model that predicted the relationships and their directions in the group.

In the second part of their study, Maduelli and Wing used force transducers to record the movements of two participants. One participant was designated to be the leader, was clapping his/her hands, while the other participant was the follower not audibly clapping (to prevent bi-directional coupling between the participants) but rather squeezing two force sensors, one in each hand. They experimented with leading and following changes in the timing and force of actions. Looking at timing, they observed lag 0 cross-correlations within the follower’s hands, while the cross-correlation between the leader and the follower had a peak at lag 1, meaning that the

follower was following the leader's tempo changes with a one beat lag, but follower's both hands were adjusting at the same time, without lag in between them. This correspondence in timing was present also when the task was to imitate changes in force, while there was no correspondence in force while timing was in focus. Timing turned out to be the more salient cue of the two (Maduell & Wing, 2007).

Even larger groups of performers were studied by Lucas, Clayton, and Leante (2011). Their focus was on inter-group entrainment in Brazilian Congado. Congado is a religious tradition in the Minas Gerais region of Brazil, where different groups celebrate a religious festival by parading around the town in groups of up to 50 musicians and dancers. Different groups, identified with one of three communities, Congo, Moçambique or Candombe, all contribute to the overall celebration, but have their own distinct parts and identities. Each group consists of a lead singer, three drummers and three to six percussionists and a large group (up to 40) of dancers; they have their own musical repertoire, distinct costumes, banners etc. The groups often hear each other and even physically meet during these parades. When they meet, they perform a greeting ritual, where they try to resist entraining to the other group, as keeping their own rhythmic identity is very important. Based on Lucas's extensive ethnomusicological fieldwork, she and her colleagues decided to investigate the between-groups resisting of entrainment systematically. They recorded audio and video using multiple camera crews and drum-mounted microphones from four of these close encounters of groups and analysed the temporal relationships of their rhythms.

In these four encounters, they witnessed a range of phase and period relationships between the groups. In some instances, the groups entrained and phase-locked so that they had a zero period and phase difference. In another, interesting case, they entrained but maintained a 223 degree phase difference. In this case, both group played a three-beat pattern. For both groups, the three beat patterns were slightly uneven, so that the ratios of the durations were not exactly 1:1:1. Also, for both groups, the patterns were uneven in slightly different ways, but the patterns were stable. The way these two different three-beat patterns were coordinated was that one of the three beats was synchronised while the other two were not. In some cases the groups managed to resist entrainment, mainly by shifting their tempi to opposite directions. Lucas et al. identify three factors that contributed

to the groups entraining or resisting it. The first was the strength of their coupling, measured as their physical proximity (influencing the loudness of the other groups' playing in relation to one's own) and whether the groups have visual contact to each other (influencing the possibility to entrain to the other bands' visual cues). The second factor was the tempo difference between groups and stability of the tempo of the given rhythm for the group. A group might have had a higher tendency to speed up while playing certain rhythms than others—playing that rhythm while meeting another group playing in a faster tempo resulted in period convergence and eventually phase locking. Finally, the group's intention made a difference. Lucas et al. note that if two groups of the same community (e.g. Congo or Moçambique) met, they did not try to resist entrainment, unlike when meeting a group from the other community. In the latter case, entrainment did occur but much more intermittently and in a less detectable phase relationship (e.g. the 223 degree pattern). In fact, participants themselves did not realise having entrained in that case. They considered this non-zero phase relationship as independent enough, yet easier to maintain than a non-zero period difference. (Lucas et al., 2011.)

Mutuality of adaptation was even observed in a rowing eight (Wing & Woodburn, 1995), where the assumption was that the stroke (the rower closest to the stern of the boat, the one who effectively sets the pace) influences the rowers behind him/her (rowers towards the bow of the boat), while the influence would not extend to the other direction, as crew members can only see those sitting in front of them. However, they did not observe such asymmetry in the force dynamics, which indicated that rather than the crew synchronising in a chain so that number 7 syncs with the stroke, number 6 with number 7, et cetera until bow syncs with number 2, the whole crew entrains to a common rhythm, manifested in the movement of the boat. Indeed, based on personal experience of rowing for the St. Edmund's College Boat Club, the entrainment within the crew is a holistic matter and while the stroke has special responsibility for maintaining the rhythm, and is the person who the rest of the crew needs to follow in changes of stroke rate, both the intention and the experience is that of entraining together as a crew. This can be felt both as a stroke and in any seat. Coupling through the movements of the boat is very strong and visceral, but it is strengthened also via auditory cues coming from the blades, as well as visual cues from

the movements of the boat, the blades and the other rowers. A crew can not function effectively unless all eight rowers independently have a good concept of rhythm. This corresponds with the criterion of oscillator independence: there is no room for resonators in the crew. Sometimes an independent oscillator runs out of fuel and turns into a resonator - such incident was observed in the 2002 Boat Race, when Cambridge oarsman Sebastian Mayer collapsed of exhaustion, and while he tried to keep with the rest of the crew, he no longer was an equal and independent part of the crew and Oxford managed to overtake Cambridge and stormed to a relatively easy victory.

4.4 Conductor studies

The conductor works to keep the orchestra synchronised and in time, as well as directing the expressive features such as tempo variations and dynamics of the orchestra. Of course the bulk of the conductor's work takes place at the rehearsals, where conductors shape the orchestra's performance of the piece according to their artistic vision and taste. Thus their leadership is mostly exerted before the actual performance, but there is also the more narrow question of how the leading and following occurs in terms of timing of the conductor's gestures and the playing of the orchestra.

First of all, as the orchestra tries to synchronise, the conductor is not the only source of timing information. In fact, the other musicians in the ensemble have a more important role in providing timing information for the performer (A. Clayton 1986; reported in Luck & Nte, 2008). Clayton investigated the role of different sources of timing information on ensemble performance (other players, conductor, score, performer's own sense of rhythm), and concluded that of these, the other players and the conductor were the most important ones, in this order. Indeed, smaller ensembles operate without a conductor and even in larger orchestras the role of the conductor in terms of keeping the timing has been proposed to be a "general" one, while the detailed timing is achieved by listening to one's peers (ibid.).

Clayton's findings suggest that the coupling between the conductor and the orchestra is weaker than the coupling between performers. Conductor's beats are visual, and rhythmic movements are not attracted to visual beats as strongly as to auditory ones (Repp & Penel, 2004). The clarity of con-

ductor’s beats depend on the conductor, as well. In Geoff Luck’s studies, it has been shown that the timing of the conductor’s beats is perceived at the point of maximum acceleration along the trajectory, not at the lowest position point of the gesture, as one would intuitively think (Luck & Nte, 2008; Luck & Toiviainen, 2006). Also, it has been observed that the orchestra can be radically out of phase with the conductor, yet still be perfectly in sync within. Anecdotal stories about tuba players pulling the whole orchestra out of sync with the conductor only to see how they react also demonstrate the effect that was experimentally shown by Repp and Penel. The match between the rhythmic blasts of the tuba and the equally rhythmic but silent waves of the conductor would not be an even one.

4.5 Leading and following

Interactions in an ensemble are about much more than just timing. The issue of roles, for example, comes up naturally when talking about interaction. In many everyday interactions and networks, people have different, often conflicting or complementary roles. One simple way to categorise roles, is to group them to *leaders* and *followers*. This is also a prominent role categorisation in music, where conductors or soloists may assume leadership roles over orchestras or accompanists. This is true also for chamber music ensembles, where there are different leadership roles in both social and musical domains (Davidson & Good, 2002; Goodman, 2003; King & Ginsborg, 2011; Murningham & Conlon, 1991). But how, if at all, are these roles implemented in the actual musical interaction?

As was mentioned above, Rasch (1979, 1988) found that the leading voice in a trio is typically ahead of the other voices. On the other hand, in a simpler context of a tapping experiment, synchronising with a metronome, the direction of influence is clear—the human participant follows the leader, the metronome. And yet, in a typical synchronisation trial, we observe so-called “negative mean asynchrony”, which means that the participant taps, on average, ahead of the metronome by about 30–50 milliseconds (Aschersleben & Prinz, 1995). While this one-sided case is perhaps not a good model for actual interaction, it suggests that determining leadership based on the advance of the onsets alone is not a reliable method. Luckily, this issue has been investigated in a couple of studies directly, with a more complex view

emerging from those.

For example, Friberg and Sundstrom (2002) showed that in jazz band improvisation, modifying the amount of asynchrony and being ahead or behind the beat is done in effort to achieve “swing”. Specifically, the soloist was often playing “behind the beat” especially on the downbeats (by 30–90ms in slower tempi, less in fast tempi), while being synchronous on the offbeats. This relationship, as well as the “swing ratio”, or the ratio of durations of consecutive quavers both vary depending on the tempo of the piece.

Social and musical interaction are intertwined in the performance, not to mention in the practice period leading up to the performances. There is a growing body of research on the practice process, indicating that the musical and social interaction in the ensembles go hand in hand (Davidson, 2009; Davidson & Good, 2002; Murningham & Conlon, 1991). Interaction in the ensembles is a mixture of verbal, non-verbal and musical communication. Williamon and Davidson noticed that majority of practice time (about 90%) in a piano duet was spent on playing (Williamon & Davidson, 2002), and often playing was a more efficient way of solving performance-related problems than talking about them (Davidson, 2009, 370). During practice, coordination and communication works towards establishing a unified representation of the musical work to be performed, but also the ensemble needs to develop “shared performance cues” that they can use to coordinate their performance (Ginsborg, Chaffin, & Nicholson, 2006). In this process of building a musical understanding, also the movement styles of performers converge as they adopt each others’ movement styles (Williamon & Davidson, 2002). Williamon’s and Davidson’s study was based on observations made from video recordings, but this question could be studied in a more exact way using motion capture, which can quantify movement at a high accuracy. In the literature on social and musical interaction in ensembles it is clear that while there might be different social roles, most of the negotiation and decision-making is done via musical interaction, relying on the mutual adaptation process to achieve a common musical goal.

There are not many studies that have examined entrainment in dyads or groups in actual music-making situations. Goebel and Palmer (2009) investigated dyads of pianists, whom they assigned to leader and follower roles. 16 pianists were recruited. In the study they were playing tailor-made duets on the same piano, each using only their right hand, while their playing was

recorded and their body movements recorded using a motion capture system. Goebel and Palmer manipulated role assignments and auditory feedback conditions so that they could look into the leadership dynamics in the dyads.

Their results show clearly how entrainment in a dyad is a result of an ongoing, mutual adaptation of timing. When the pair could hear each other, positive and highly significant cross-correlations both at lag -1 (leader lagging) and lag +1 (follower lagging). The authors explain this with the participants' error correction processes being in anti-phase, but then also suggest that it is due to "two-way tracking processes", which essentially means mutual adaptation to each other. This mutuality was there even though one participant was assigned the role of a leader, and s/he was playing the solo part (higher register/usual right hand part). The only way Goebel and Palmer managed to break the mutuality was to disrupt the coupling in the pair by cutting off the auditory feedback from one participant to the other. These results are very similar to what Konvalinka et al. (2010) found in a finger tapping experiment (see section 4.2), and also similar to my results in pilot studies. In cross-correlations, both positive and negative lags are significant, indicating that regardless of the instructions, adaptation is mutual whenever it possibly can be, and to disrupt this is the only way to introduce unidirectional tracking or influence (I will return to this issue later when discussing cross-correlation, in section 6.5).

So far all duetting experiments have had one aspect in common: leader and follower roles could be assigned, but in general they did not influence the mutuality of entrainment. Regardless of who is supposed to lead or to follow, unless you simultaneously also disrupt the coupling in the dyad, synchronisation will be achieved and maintained in a process of continuous, mutual adaptation. This works in the realm of predictable and stable musical beat, of course, as disrupting the beat completely will make this type of anticipatory process virtually impossible. In the absence of predictable shared representation, interaction can turn into a reactive affair where the delays between actions and reactions grows by at least a couple of orders of magnitude compared to the millisecond accuracy of the predictive processes.

A good demonstration of the "power of mutuality" came from the study of Lior Noy et al. (2011), who studied the "mirror game", an improvisation theatre staple where two people stand facing each other, mirroring each other's. This warm-up exercise often starts with one participant being the leader that

the other follows, as if one participant is standing in front of a mirror, making all kinds of movements, that the other participant, being the reflection, tries to mirror in real time. Eventually the game progresses to joint leadership, where the two participants create these movements together. In their study, Lior et al. simplified this full-body mimicry game to a movement along just one spatial dimension. They made a table-top box that had two handles that could be moved side-to-side along parallel grooves. Each participant would move one of the handles, and they would do so in two conditions: either one participant would be the leader that the other was supposed to follow, or they shared the leadership. Their results are interesting: participants were often moving in better synchrony when they shared leadership. Noy et al. call this “co-confident motion”, which is a lagless, jitterless, and synchronised motion. When one participant was leading, the follower would often not be able to follow immediately when the leader changed direction, then would overshoot when attempting to catch the leader etc. Similar results have been reported in a finger-pointing task by Yun, Watanabe, and Shimojo (2012), showing that continuous, body-movement entrainment also can occur in a context that is not rhythmic or beat-based like music or dance.

Is leading and following just a role that you can take or is it a personal characteristic? This was investigated by Nadine Pecenka and Peter Keller (2011). As has been established earlier, synchronisation to a metronome is an anticipatory process, where the next metronome onset time is predicted so that a motor response can be timed to coincide with it (Aschersleben & Prinz, 1995). Some people are better than others in sensorimotor synchronisation tasks, and music training seems to help (Pressing & Jolley-Rogers, 1997). Pecenka & Keller wanted to test whether people differ in their ability to anticipate changes in metronome. To this end they devised a test where the metronome tempo fluctuates continuously, following a sigmoidal function. They found that some participants were indeed able to learn to predict these changes and anticipate them when tapping. Others would be more reactive, and change their tempo only in response to the metronome, with a slight delay. In the second part of the study, they formed pairs for a cooperative tapping study, mixing the predictive and reactive tappers.

By testing people in consecutive sessions, Pecenka & Keller could establish that these tendencies to predict or react were in fact stable. In the cooperative tapping studies they found that pairs consisting of two predictive

tappers outperformed mixed pairs and those that consisted of two reactive tappers. The ability to track tempo changes in a predictive fashion is thus a skill that is relatively stable, and at least to some degree independent of musical expertise (most participants in this study, both the high and low predictors, were amateur or professional musicians). This predictive tendency also has a measurable effect on interaction, benefiting both the stability and entrainment accuracy in the cooperative tapping task. While anecdotally all musicians know that some people are better ensemble players than others, we still know relatively little about what these ensemble skills exactly are, how they could be measured, and how they can be improved or taught. This is an area where focusing on the individual has had a negative effect on applicability of the research results.

A more embodied approach was taken by Moore and Chen (2010) in their study on entrainment of two members of the string quartet. Instead of looking at the notes played, they recorded the right forearm (bow hand) movements with angular velocity sensors to get time-series data of the bowing motions. Their analysis is methodologically very interesting. For example, they showed that notes are not produced in a fixed tempo, but instead, interactive coupling (mutual adaptation) adjusts the timings helping to keep asynchronies in the note segments in check. To demonstrate this, they picked two ten-second samples of the data, where the up-bowing and down-bowing are in very good sync (peak velocities are somewhat different but the shapes and most importantly the zero-crossings are simultaneous). Then, they compared viola in the first segment with the violin in the second segment, showing that the resulting pair of curves were not synchronous, as their zero-crossings were out of phase. This showed that the two processes were not perfectly stationary¹ and fixed in tempo but were “living” and interacting processes.

Moore and Chen also note that performers seemed to generate the patterns of four semiquavers in a consistent fashion—and their timing patterns could be modelled using just 4 histograms of up- and down-bow intervals and convolution, or as an alternating renewal process. This supports the idea that instead of focusing on individual notes, musicians chunk them to

¹A stationary process is one whose parameters such as mean or variance do not change over time, so for example it does not drift or have sudden changes of its properties (see e.g. Shumway, 1988).

larger structures (motifs, riffs or licks) and then attention is directed to these higher-order structures while the lower (individual note) level is subconsciously steered by over-learned motor programmes (Palmer & Meyer, 2000).

Moore and Chen also investigated the synchronisation in the pair comparing correlations of original data to correlations of shuffled data. This shuffling should remove all links of direct influence, and just leave the pattern within the note group that didn't depend on interaction. The sequence of synchronisation errors was massively changed by shuffling (errors grew larger than the average period). These evidence suggested that "[s]ynchrony was maintained by some form of interaction whose effects were removed by shuffling" (Moore & Chen, 2010, 413).

4.6 Summary: entrainment and leadership in dyads and groups

Conducting rhythm and timing experiments in dyadic or group settings is a relatively recent trend, though the motivation for doing these studies has been there for long, and some pioneering work was conducted in the 1970's and 1980's (Rasch, 1979, 1988; Shaffer, 1984). This is partly due to development of methods, such as MIDI and especially digital recording and sound analysis technology, as well as various motion sensors (see appendix B for more details on data collection methods). Rasch had to use an anechoic chamber, special microphones and a very arduous analysis of waveforms (essentially by hand), just to extract onset times from his trio recordings. The same analysis can be done easily today using contact microphones, accelerometers, or optical motion capture to collect the data, and automated algorithms to extract all kinds of acoustic and kinematic features from it.

Another reason for this trend are shifts in epistemology. More and more researchers are taking the embodied, situated and social agendas seriously, and are developing alternatives to testing isolated individuals in laboratory settings. The importance of naturalistic settings and studying actual interaction have been recognised. Also, after more than a hundred years of rhythm and timing experimentation in individual settings, it is getting more and more difficult to come up with novel experiments. Actual interaction is dynamic and as such can produce surprising results. Having two people inter-

acting is qualitatively and phenomenologically different from the individual setting, and interesting things emerge from the dynamic process of mutual sense-making and behavioural adjustment. Now that some of these effects have been demonstrated (e.g. Himberg 2006; Himberg and Cross 2004 and the various studies discussed in this chapter), and the major epistemological and methodological barriers no longer exist, the interest in studying these phenomena keeps growing.

The issue of leadership in timing behaviours is a complicated one. As was established above, simply being the first one to tap is not an indication of leadership. Also, our tendency to mutually adapt to each other overrides any instructions to assume roles of leaders or followers, making this a difficult issue to study experimentally. Maduell and Wing propose in their paper (2007) that leader would be the one who controls the performance, and they illustrate this (their figure 7, p. 616) by showing that the leader initiates tempo changes, to which the follower then responds at a lag of one beat. While this is a good way to illustrate the adaptation process that cross-correlation analysis is sensitive to, this illustration ignores another important aspect. As we demonstrate in our African dance -study (chapter 8), and has been discussed elsewhere (e.g. Pressing & Jolley-Rogers, 1997), experts are more steady in their timing than novices, both in maintaining their tempo as well as in beat-to-beat stability. From this perspective, it could be argued that leader is someone who is stable and steady, makes fewer changes and corrections, is somewhat metronome-like. Maduell & Wing noticed that some control cues were used to maintaining status quo, so the leader can explicitly lead by indicating that no changes are needed or desired to the current performance.

Music is often about *both* stability *and* change. The role of the leader is to focus on whichever is required by the composer, the collective improvisational idea, and/or the conventions of the genre. The role of the leader often becomes visible at pivotal points, beginnings and ends of movements or changes of e.g. tempo, or even outside the performances: in deciding who to include in the group, what to play, how to play it, etc. Leaders might have special responsibilities during the performance, in playing the solos or conducting the performance, but the performance itself emerges from mutuality of adaptation and seamless collaboration. Certainly the issue of leadership is more complex than just being ahead or behind in time.

Hence, other kinds of analyses of the direction of influence are needed. One approach is to look into the meaning of gestures, taking a qualitative observational approach as did Davidson & Good (2002) or Maduell & Wing (2007). However, there are also methods that can be used to either visualise, quantify or even statistically evaluate *influence* in dyads or in groups. These methods go beyond looking at the tightness of entrainment or other aspects of timing in the dyad or a group. In the following sections I will discuss a few of these methods, for example windowed cross-correlation and Granger causality, and demonstrate how they can be applied in studying musical interaction. Regardless of which methods are applied, it is important to know what leadership *means* in the musical context in question. Therefore a thorough ethnographic analysis, or qualitative observational analysis is often needed before using these “mechanistic” quantitative analyses. The studies by Maduell & Wing (2007) and Lucas et al. (2011) are good examples of this kind of multidisciplinary approach.

Chapter 5

Phase relationships: circular statistics and cross-recurrence

Looking at relationships between multiple dynamic processes, a usual method is to investigate their phase and how the phases of the processes are related to each other (Pikovsky et al., 2001). Haken, Kelso & Bunz (1985) studied phase relationships and synchronisation, and constructed the classic HKB-model of phase transitions using a “finger-wiggling” task. They observed that the system made an abrupt transition from anti-phase to in-phase as frequency was increased.

The HKB model describes the dynamics of a system, where the phase relationship of two oscillators ϕ has two stable coordination models, in-phase ($\phi = 0$) and anti-phase ($\phi = \pi$). In this model, the frequency of oscillation (tempo) was found to be the control parameter, as in low frequencies the systems were found to be bi-stable (both coordination modes, in-phase and anti-phase were stable), but with higher frequencies, the anti-phase system experienced a transition to in-phase, whereas in in-phase systems no transitions were observed. Later, other researchers explored hand turning and other bimanual movements as well as dyadic (two-person) coordination tasks. In these tasks the HKB model has been seen to fit to dyadic interaction (R. Schmidt, Bienvenu, Fitzpatrick, & Amazeen, 1998), and it has also been extended to include in-phase to anti-phase transitions as observed in hand pronation and supination tasks (Carson, Goodman, Kelso, & Elliott, 1995; A. Fuchs & Jirsa, 2000).

In this chapter I will present two approaches for looking at phase rela-

tionships. First, I look at a collection of methods called circular statistics. These are tools for analysing angular data, especially periodic and discrete data, such as tap onset times in a rhythm tapping study. The idea is to wrap the time-series around a circle, so that each tap is measured not by its temporal distance from the beginning of the trial, but in relation to the underlying period. This has many advantages, to both data visualisation and analysis, as will be discussed below.

The second set of tools is called cross-recurrence analysis (CRA). These methods are for situations where the two processes are not necessarily periodic, might even be noisy, and measurements are continuous, as in the case of body movement data. These tools can be used for detecting entrainment and influence even in situations where the relationship between the participants is not as clear as in musical interaction. These methods have been used in looking at behavioural synchrony in conversations, for example. They too are based on phase relationships, but here phase is defined more generally, as the processes are not periodic. Instead, we reconstruct the phase spaces of the processes we want to analyse and then check if the two processes travel along similar trajectories through their phase spaces.

I will present the main ideas and methods in this chapter. For more details, see appendix C.

5.1 Circular statistics

Circular statistics (Fisher, 1993; Jammalamadaka & Sengupta, 2001) are a set of tools for analysing angle or orientation data sets. There are examples of this type of data in many fields of science from geography to medicine, and from engineering to music. In most of these fields, the collected data is originally measured in angles, like for instance the direction of currents in oceanography, the direction of fault lines in geology or departure angles of migratory birds in biology. In addition to these cases, this methodology suits all data that can be *converted* into angles, for example data that is periodic, or represents a periodic process, such as tapping data. Onset times collected in a tapping experiment are typically analysed as time series data (Shumway, 1988), but can be converted to phase angles and analysed using circular statistics. This approach is advantageous especially when the focus of research is in beat-by-beat timing accuracy and synchronisation,

and not so much in the large-scale tempo drifts or other trends, for which time-series analysis is the better option.

Circular and time-series analysis are complementary to each other. Time-series analysis is handy for investigating drifts and trends, as well as preliminary investigation of periodicities in the data (e.g. autocorrelation), while circular methods provide flexible tools for analysing consistency and synchronisation (e.g. relationships between multiple tappers). Missing data poses fewer problems than for linear analyses, and circularity can remove the effects of tempo drifts. Circular analysis extends to complex patterns of entrainment (e.g. a 2:3 polyrhythm, where every other tap of one participant coincides with every third tap of the other). Circular statistics can be used with very simple tapping data, as there is an opposite number for all of the usual linear measures, but their biggest benefit is that they can be extended to more complicated situations, relationships and to more challenging data.

One advantage of circular statistics is that the representation of data in phase angles/vectors makes it possible to use analogous methods and metrics that one would use for analysing continuous data (such as movement data) and behaviour of nonlinear systems. Thus it allows a range of synchronisation behaviours to be studied, using discrete as well as continuous data, while maintaining a more coherent picture of the phenomena and comparability between data sets and analysis results.

In this chapter, I will first introduce circular data and some useful metrics such as \bar{R} that can be used as an entrainment measure. Then I will discuss various methods for converting tapping data from the linear, time-series domain (onset times) to the circular domain (phase angles). Finally, I will demonstrate how circular data can be visualised and applied for co-operative tapping data. This section is based on measures and methods presented in Fisher (1993). More examples can be found in the following chapters that present empirical experiments, as well as in appendix C. Some further examples of studies where these methods can be used are the pilot studies described in the appendix A.

5.1.1 Circular data

To illustrate the differences between linear and circular domains, consider the example of patterns of customer visits to a 24h corner shop. To study how the visits are distributed in time, data on visits could be collected for

one week by recording the time of each customer’s visit. Were we to analyse this data in the linear domain, we could start the clock when the recording period starts, and then tally all the times of the clock when customers entered the shop, and keep doing this for a week. The clock would run from zero to 604800 seconds, and each customer entry would have a timestamp with a value between 00:00:00 and 168:00:00 (hhh:mm:ss), telling how much after the start of the observation period this customer entered the shop. This is equivalent to the series of onset times in a tapping experiment.

The circular way of time-stamping the shop entries would be to use the time of day instead. As the linear time-series keeps growing, the “circular time” will reset every midnight: after the first 24 hours or 86400 seconds, the linear clock will continue to 86401 seconds (24:00:01) while the circular clock will start over from 1 (00:00:01). Converting times from the linear to the circular domain, we need to know the period of repetition (T) in the data (in our shop example it would be 24 hours), and take modulo T of the onset times. Thus, all onset time values are now between $0 - T$. This is called *wrapping*, as the linear distribution is wrapped around a circle, representing the period (e.g. a clock face).

A major advantage of this is that now meaningful means can be calculated. For raw onset time data, no mean can be calculated, due to a dependency between values and their order in the time series. Another advantage is avoiding the so-called cross-over issue, or bias that stems from bad selection of zero point for representing periodic data in the linear domain. For example, if we want to analyse when the corner shop is busy, we might plot a linear histogram of customer entry times, divided into 24 one-hour bins. However, the interpretation of the graph would depend on which bin we would select to start from. Midnight would be a natural choice. In this case, entries occurring just after midnight at 00:05 and just before it at 23:55 would fall into different bins at the extreme ends of the histogram. Phenomenally, however, these two customers were visiting the shop very close to each other. In a circular distribution, these would also be plotted close to each other, regardless of what the selected zero point would be.

Circular data are a series of data points in a coordinate system. The points (x, y) in the Cartesian system can be seen as vectors from the origin $(0, 0)$, thus having a certain length and direction. These data points can thus be described also using the polar coordinate system, where the fixed point

of origin is called the *pole*, and the ray from this pole, the fixed reference direction as the *polar axis*. The data points in this coordinate system are lines which can be described by the length of their radius from the pole and the angle (azimuth) between the radius and the polar axis (r, θ) .

For convenience, we set the length of each data point as 1, so that all data points fall on the unit circle. By convention (this is derived from trigonometry), the zero-phase point, or the polar axis, is set to be at “three o’clock”, and time progresses anti-clockwise. Vector $(1, 0)$ represents this zero line. In the “phasor view”, the process starts from this starting phase and then rotates around the circle anti-clockwise at a constant speed until it returns to the original state exactly one period later, having covered 360° or 2π of phase.

Let this constantly circling point represent a metronome. At constant time intervals, it reaches $(1, 0)$ and the phase resets, and the metronome clicks. Now, if someone is tapping along this metronome, we can look at where the metronome point is on the unit circle when the taps occur ¹. Taps in a synchronisation trial can now be represented simply by their phase angles: the angle θ in figure 5.1 represents the time difference between the metronome and the tap; the angles are essentially proportions of the period of the referent.

5.1.2 Mean direction and concentration measure \bar{R}

How to calculate the average direction, if our sample of taps have phase angles of, say, 1, 5, 358, 350 degrees? Clearly the arithmetic mean (178.5) is not an acceptable answer. This is a manifestation of the cross-over problem mentioned above. Instead, the preferred method is to treat the angles as vectors, and use vector addition to obtain the mean direction.

The quantity R (see figure 5.2) is the resultant length of this sum vector associated with the mean direction $\bar{\theta}$. The mean direction $\bar{\theta}$ is related to the average asynchrony. In perfect in-phase synchrony $\bar{\theta} = 0$; in anti-phase synchrony $\bar{\theta} = \pi$; tapping with a metronome often results in negative mean asynchrony, where $\bar{\theta}$ is just below 2π .

Dividing the resultant length with the number of unit vectors it repre-

¹This is called stroboscopic observation, and this is discussed in more detail below, in the section on phase conversion (5.1.5) Converting timing data from linear to the circular domain turns out to be a crucial step.

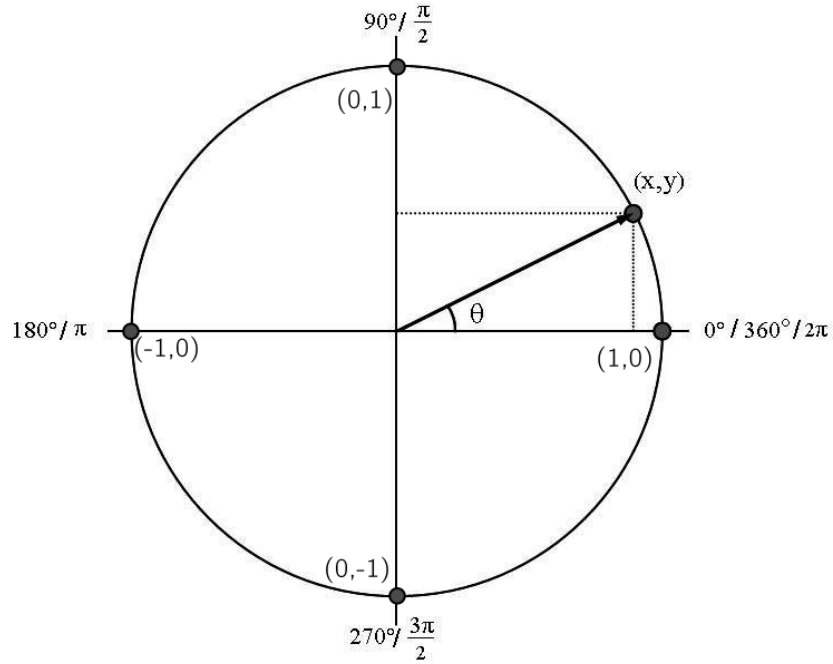


Figure 5.1: Tap onset as a point on a unit circle

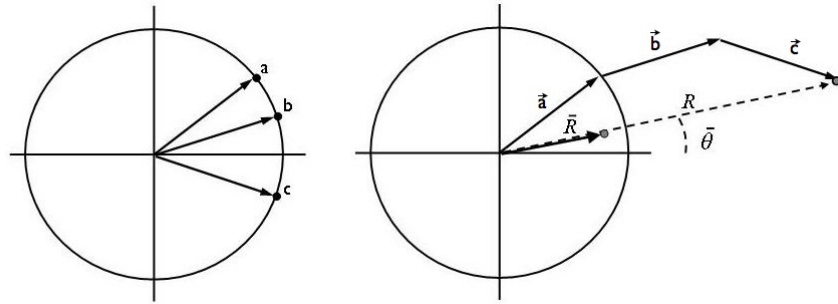


Figure 5.2: Calculating the average angle $\bar{\theta}$ and length of the resultant vector \bar{R} . Points a, b, and c on the unit circle are represented by vectors **a**, **b**, and **c**. These are joined end-to-end to calculate their sum, and the length is then divided by the number of vectors to get the vector mean.

sents gives us \bar{R} , or the *mean resultant length*. As we are using the unit circle, this value varies between 0 and 1, 1 being the case of perfectly con-

sistent metronome, producing a set of data where all points coincide: in this case all the unit vectors point to the same direction, contributing all of their length towards the mean direction. In other words, \bar{R} is a concentration measure of the circular distribution, and as such, it can serve as a synchronisation or entrainment measure, with high values indicating high entrainment. Unfortunately, the other end of this scale is not well-defined, as both a uniform distribution (no synchrony) and multimodal distributions (complex synchrony) yield values close to zero, as clusters of vectors pointing at opposite directions cancel each other out. These two cases can easily be distinguished by plotting the circular distribution, or the rose histogram.

5.1.3 Rose histogram

Florence Nightingale was the first person to use the rose plots that are now the default option of presenting histograms of circular data (Fisher 1993, p.5). She demonstrated with what she called coxcombs, how improvements in sanitation managed to prevent deaths in the military during long campaigns, such as the Crimean war. She plotted the number of deaths attributed to different causes around a circle that represented a year. With the circular presentation, she avoided the cross-over problem discussed above.

The principle of the rose plot or rose histogram is the same as that of the linear histogram: to see how the data is distributed, it shows the proportional frequencies of the data in predefined bins (see figure 5.3). The circumference of the circle is divided to n sectors or bins, that correspond to bars in linear histograms. The radii of these sectors are then proportional to the relative frequency of the data within each sector. The larger the “petal”, the more data within that bin, the more taps have landed on that sector, relative to the referent that marks the zero.

Inspecting the data visually is an important part of any statistical analysis, and circular statistics are no exception. Rose plots are a quick way to check whether the data is unimodally, multimodally or uniformly distributed, as this influences then the choice of analysis methods, and for example helps in detecting possible cases of polyrhythmic synchronisation. The most important descriptives, \bar{R} and $\bar{\theta}$ can often be estimated relatively accurately by examining the rose plots. This can be useful in for instance comparing the performance of two tappers against a common metronome, or the mutually abstracted pulse, for instance.

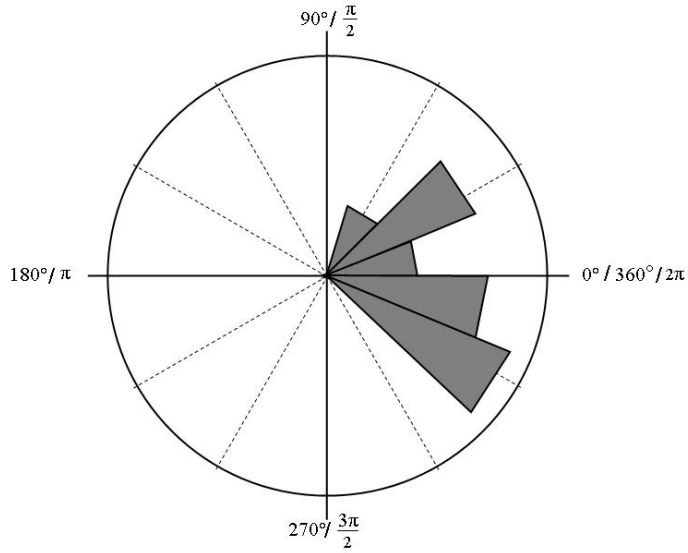


Figure 5.3: Rose histogram

5.1.4 Polar scatter plots

As a tapping trial can be summarised in two descriptive statistics, the concentration of the distribution \bar{R} , and the mean direction $\bar{\theta}$, we can summarise multiple trials by plotting these values for multiple trials into a same figure. The two descriptives determine a point in the polar graph. The value of \bar{R} determines how far from the pole the dot is (radius), and the value of $\bar{\theta}$ the direction in relation to the polar axis (angle). Figure 5.4 shows an example of a number of tapping trials plotted into the same picture. Each dot represents one trial, the phase angle represents θ and the distance from the origin represents \bar{R} . Blue dots represent trials between two human participants, green dots are trials between humans and deadpan computer tappers, and red ones are trials between humans and phase-variant computer tappers. This experiment is reported in chapter 7.

This can be used in a similar fashion as scatter plots usually are, to find relationships and structures in the data. Just as the rose plot is a handy tool for investigating individual trials, the polar scatter plot is a good way to

visualise a larger number of trials, even complete experiments. Colour-coding the dots by condition, for instance, allows one to easily compare the effect of the condition, to both the direction and concentration of the distribution.

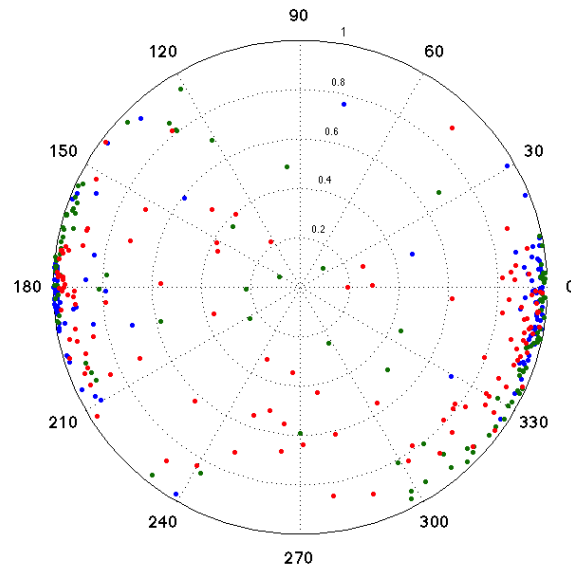


Figure 5.4: Polar scatterplot.

5.1.5 Phase conversion

General case

Converting onset time data to angles/phase values is the key stage of the analysis. Conceptually the conversion is simple but as is usually the case, the devil lurks in the details. Especially the question of referent is pivotal: in relation to *what* is the relative phase of the tap calculated? There are a few different options, depending on what aspect of the relationship between the two processes (metronome and a tapper, or two tappers for example) is of interest. So far, I have discussed discrete data, such as tapping data with discrete tap onsets. Circular methods can be used with continuous data (such as movement data), as well. With continuous data, one can use the Hilbert transform or the Hilbert-Huang transform (Huang et al., 1998; Khvedelidze, 2002) to calculate the instantaneous phase, and then use circular methods

for looking at entrainment, for example. For discrete data, one may use the *stroboscopic observation method* (Pikovsky et al., 2001). There are different ways of doing this, depending on what is chosen as the referent that defines the period, the circle.

The general case of this was already presented above: the referent, for example a metronome, defines the circle, with its onsets occurring at the zero point and the phase growing linearly between these onsets, from zero to 2π . Stroboscopic observation means that we observe the phase of this referent not continuously, but only at those time points when the participant taps their drum. Formally, the phase of a tap, occurring at a time t , during a metronome period $T_k < t < T_{k+1}$ can be calculated using the linear interpolation presented in equation 5.1:

$$\phi(t) = 2\pi k + 2\pi \frac{t - T_k}{T_{k+1} - T_k} \quad (5.1)$$

In a synchronisation trial, taps would ideally cluster into a very concentrated distribution near zero. The angle of the peak of this distribution, $\bar{\theta}$, would indicate the average phase difference (synchronisation error) of the trial. Peak just above zero would mean the tapper is in average a bit late, while values just below the zero line would mean she is a little bit ahead of the metronome, in average. The latter is often the case in finger tapping, in the linear domain it is called negative mean asynchrony (Dunlap, 1910, Aschersleben & Prinz, 1995).

Syncopation, or tapping in between the metronome beats, would place the peak at the opposite side of the circle, π out of phase with the metronome. This relationship is often called anti-phase synchrony, as it is a phase relationship that is relatively stable (easy to maintain) (Kelso, 1995).

Looking at tapper's onsets in relation to a metronome, the concentration measure \bar{R} reflects the synchronisation accuracy, or phase locking of the tapper, the higher the value, the better the phase locking.

Self as referent

Metronomes are handy referents, but usually musicians need to keep time and maintain a steady beat without the help of an external referent. In Stevens' classic experiment (Stevens, 1886), these two aspects of time control (synchronising accurately and keeping a steady tempo) were combined into what

later became the staple of rhythm and timing studies, the synchronisation-continuation paradigm (Repp, 2005a).

At the synchronisation stage, metronome serves as the target for the tapper, and as a referent in the analysis. The referent does not need to be “real”, however, but it can be simulated. For example, we can analyse the tapper’s taps in relation to where the metronome clicks *would have* occurred, had the metronome continued. This can be interesting, if we want to know how well the original tempo is maintained. This measure is very sensitive to even small perturbations, not to mention more systematic drifts in tempo, however.

This sensitivity can be seen from the example in figure 5.5. By changing the referent period by only 0.05 seconds (from 0.5 to 0.55 seconds), the distribution changes drastically and very quickly.

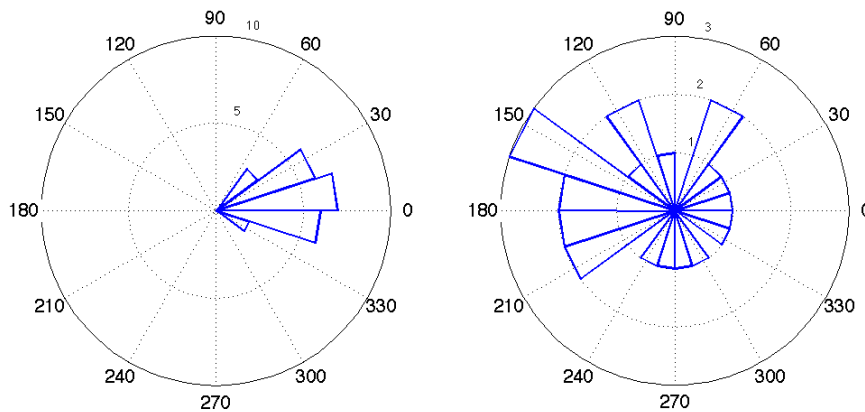


Figure 5.5: Comparison of rose histograms calculated from the same data with different referent periods

A better option would be to measure the tapper’s performance against his/her own performance. For example, the average period of the participant could be the referent, either the global average or one calculated in a moving window. Alternatively, following what Guy Madison has suggested in the linear domain as a measure of tapping consistency (Madison, 2004), the participant’s previous period could serve as a referent. Madison’s measure of variability is based on calculating the average amount each inter-tap interval

differs from the one two beats before it. Human tapping often displays negative lag 1 autocorrelation of the ITI, which means that successive inter-tap intervals are often rather dissimilar, even when maintaining a steady tempo in continuation tapping. This can be seen in the typical zigzagging ITI-plots. Thus the period / ITI two beats back provides a better comparison.

When phase conversion is done in relation to the participant’s own tapping period, the resulting distributions are often quite concentrated, with $\bar{\theta}$ close to zero. In isochronous tapping, this measure would also be rather blind to sudden tempo changes, but react better to gradual tempo drifts. A sudden change to a new tempo would only produce one data point, one phase value that is not close to zero: when the first tap in the new tempo is compared to the last one in the old tempo. After that the referent changes, and the concentration of the distribution suffers no further damage, even the peak stays put in zero. Only slow drifts or erratic taps, or a very high amplitude zig-zagging would spread the distribution. High \bar{R} values indicate high beat-to-beat similarity. Therefore the \bar{R} relative to tapper’s own previous taps as a referent can be used as a measure of *stability*.

The other tapper as the referent

In dyadic tapping experiments, the most interesting aspect and the key research question usually concerns the entrainment of the participants rather than either participants’ synchronisation with the metronome, or their individual performance (although comparisons of these three can be very interesting). Thus, the referent against which the tapper’s onsets are compared should be the other tapper.

In this conversion, the other tapper serves as the “metronome”, albeit a constantly varying one. Just as when calculating the phase in relation to the metronome, we now observe the phase of one tapper at onset times of the other. As before, the phase of the referent grows linearly from 0 to 2π during each inter-onset interval and resets to zero at each tap.

In terms of missing data, this means that the referent should be a “complete” data set without missing taps, while the “following” participant can have missed some taps. In terms of the concentration of the distribution, it makes otherwise no difference which of the two tappers serves as the referent. Comparing distribution peaks and shapes, the two possible phase error data sets (tapper 1 as referent; tapper 2 as referent) are symmetrical, the polar

axis or the zero line serving as the axis of symmetry. Therefore for most analyses, the choice of referent makes little difference, if there are only few differences in the number of missing values.

As with having the metronome as the referent, the position of the peak tells what the phase relationship between the tappers is. And similarly, a very concentrated distribution, i.e. a high \bar{R} , indicates that the two participants are entrained to each other. A more uniform distribution around the circle suggests a lack of entrainment, and thus the \bar{R} measured from the relative phase, can be used as an *entrainment measure*. Whether the distribution differs statistically from uniform (whether there is entrainment) can be tested using the Rayleigh's test, although it is not a very discriminating test. Rayleigh's tests will be discussed in detail in a later section on analytic statistics 5.1.6. Unlike the two previous phase conversion methods, tempo drifts or even highly variable inter-onset intervals will not necessarily spread the distribution around the circle, as the only factor that matters is that the two tappers are similar to each other.

Summary of phase conversion methods

In this section, a number of variations of phase conversion have been presented. In general, converting onset times in the linear domain to phase angles in the circular domain is conceptually straight-forward and also relatively easy computationally. The process is called stroboscopic observation, as the continuously and linearly growing phase of one periodic process (the referent) is observed at certain times determined by the other process (the tapper). The choice of referent is the most critical one, as the phase values recorded will greatly differ depending on that choice. Figure 5.6 demonstrates this difference.

In the figure, the blue dots represent metronome clicks, and the green and the red ones the taps of the two participants. The phase value associated with the red tapper's third tap (marked as a red star) is calculated, using the three different methods detailed above. In the linear domain, this tap lands 1.6 seconds after the start of the trial. In the circular domain, the phase value depends on the choice of the referent. If the referent is the metronome, the phase value is determined by how far in the current cycle (one cycle = 2π radians or 360 degrees) the phase of the metronome has advanced by the time of the tap. 0.6 seconds into the 1 second cycle yields the phase value

of 3.77 radians or 216 degrees ($0.6/1 * 2\pi$).

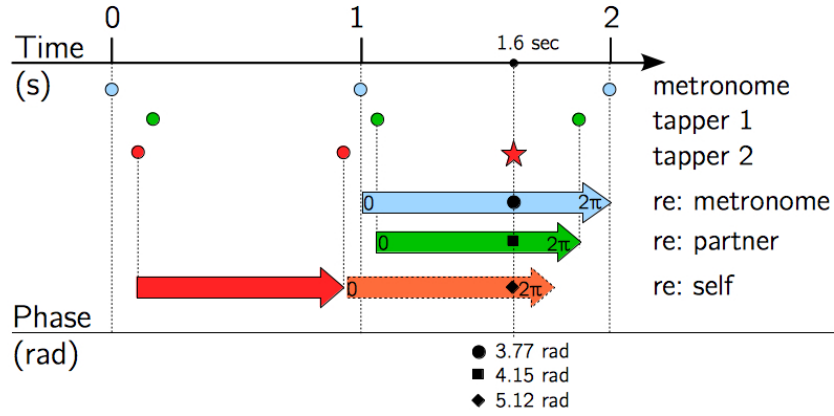


Figure 5.6: Phase conversion of a tap at 1.6 seconds (red star) using either the metronome (blue), the partner's tapping (green) or the tapper's own tapping (faded red) as the referent.

If the referent is the other tapper, as in chapter 5.1.5, then we are interested in seeing how far in its cycle the other tapper is. The current inter-onset interval of the other tapper (the green dots) is somewhat shorter than one second, only 0.82 seconds. As the previous green tap occurred 0.12 seconds after the previous metronome onset, the red star occurs when 0.54 seconds of the green ITI has elapsed. This yields us the phase value of 4.15 radians or 237 degrees ($0.54/0.82 * 2\pi$).

Finally, the conversion can be made self-referentially, looking at what the phase of red tapper would be in the case the current ITI were of equal length as the previous one was. The lighter red arrow depicts this. This time, the red star, the tap we are measuring, occurs at a time point where using the same formula as above, the phase value is 5.12 radians or . In all these cases, locally, the referent is considered periodic at 2π , which means that in the case the measured tap occurs after the referent reaches 2π , the returned value is the value modulo 2π , and thus always between 0 and 2π . The three methods illustrated in figure 5.6 are distinct, and one could argue they represent independent yet related dimensions of timing performance.

\bar{R} is the metric that represents the stability of the phase relationship be-

tween the tapper and the referent. It is a measure of concentration of the distribution of phase angles. What this stability means exactly depends on what is used as a referent (see 5.1.5)². In the case where metronome is the referent, \bar{R} measures synchronisation or *phase locking* with the metronome. When the referent is one's own taps, high \bar{R} values refer to high self-consistency or *stability*. In a dyadic case and relative phase distribution, \bar{R} serves as an *entrainment measure*.

Crucially, \bar{R} does not depend on the mean direction $\bar{\theta}$, but is only concerned whether the phase relationship stays stable. This means that \bar{R} does not distinguish between synchronisation and syncopation, which makes it a versatile and useful measure. The $\bar{\theta}$ in turn serves as a measure of the amount of asynchrony, as it quantifies the average phase difference between the tapper and the referent. These two can be used in conjunction to summarise participants' performance in a tapping trial.

5.1.6 Analytical methods

It is possible to summarise aspects of tapping trials using the metrics introduced above, such as \bar{R} and $\bar{\theta}$. Taking into account the limitations in range that these metrics have, and the potential cross-over problem in the angle values, they can be used as variables in “traditional” linear statistical models, such as regression, correlation or analyses of variance. This is mostly what I have done in the studies presented in this thesis: the key variables have been versions of the \bar{R} , and analyses have been linear. For studies where the directions or angles are important as such, there are many kinds of circular probability distributions and analytical methods that can be used with circular data. For example, the equivalent to the normal distribution in circular domain is called von Mises distribution, and it can be used for parametric statistical tests just like its linear variant. I will briefly talk about one statistical test, the Rayleigh test, and the uniform and unimodal distributions that it distinguishes³.

Circular distributions can be continuous or discrete, just as linear ones.

²I have called this same measure “coherence” before Himberg (2006), and other terms have been used by other authors, but “stability” seems to be the best description of what the metric represents. For instance “accuracy” could be confused with measures of synchronisation accuracy, such as $\bar{\theta}$.

³Books by Fisher (1993) and Jammalamadaka and Sengupta (2001) are good and thorough presentations of these methods.

Jammalamadaka & SenGupta (2001, 30) list four ways of generating circular distributions. One is *wrapping* that was already mentioned above in the section 5.1.1. There one wraps the linear data around a unit circle so that all data points that are separated by the length of the period T (the circumference of the circle, 2π), fall on the same spot on the circle. In addition to this, one can characterise the properties of the distribution, or take a stereographic projection of a linear data, or make a so-called off-setting, where bivariate polar coordinates are transformed to their directional components (ibid.).

In addition to the uniform, Jammalamadaka & SenGupta list the cardioid, triangular, circular normal (von Mises), offset normal, wrapped normal, wrapped cauchy, and general wrapped stable distributions. In addition to these univariate distributions, there are also bivariate circular distributions. Using these distributions, it is possible to define statistical methods such as correlation and regression in the circular domain. As mentioned earlier, this is outside the scope of this thesis. Nowadays, circular statistical methods have also been implemented as software packages, available for the most popular platforms. For MATLAB, there is a *CircStat* package that contains many statistical tests for unimodal distributions, as well as paired tests and measures of association (circular correlation) (Berens, 2009).

Uniform distribution U_c

If circular data are distributed uniformly (randomly) around the circle, then all directions of the vectors are equally likely. In the uniform distribution, mean direction is undefined, mean resultant length is 0 and dispersion is infinite. In terms of probabilities, the uniform distribution has a constant density $f(\theta) = \frac{1}{2\pi}$, where θ can get values between 0 and 2π .

The way to test whether a distribution of phase relations differs from uniform has been used as a crude method of detecting synchronisation between two processes. If two processes are “not synchronised”, then there is no reason to expect that any particular phase relationship is more likely than any others, and unless the two processes have exactly the same period (which can be due to being driven by the same, third process, see section 2.1), their phase relationship will eventually wrap around the circle, forming a uniform distribution. Even slight synchronisation will, however, detune the relationship and increase the likelihood of some phase relationships. This leads to a

“bump” in the uniform distribution.

5.1.7 Rayleigh test

There is an omnibus test that simply tests whether the sample distribution differs from uniform, this is called a Rayleigh test (named after Lord Rayleigh). There are two versions of this test, one tests against any unimodal distribution, without specifying a direction. The direction-specific version (also known as V-test) tests for a unimodal distribution in a specific direction. (Fisher, 1993). This specified direction means that we only test if there is a certain phase relationship between the two tappers (such as them tapping anti-phase, and consequently having a mean phase of app. π). A simple way of doing this test is to calculate the the statistic \bar{R}_0 and then compare it against critical values. The cosine of the differences between the θ values of the sample and the specified direction are calculated and the mean resultant length of this set of vectors is then calculated.

When calculating the unspecified version of the test, the \bar{R} of the sample is simply compared to a critical value, and the null hypothesis rejected if \bar{R} is too large. Fisher proposes that for relatively large samples ($n \geq 50$) the significance probability P can be estimated as $P = \exp(-Z)$ where $Z = n\bar{R}^2$. This test is quite general and can yield false positives especially with small n . The test variant that also looks if the unimodal distribution points to a specific direction has more statistical power, but even that can lead to false positives if used with small samples.

This problem stems from the generality, and the uniform distribution that the distribution is measured against: with very few data points, it is actually difficult to spread them uniformly around a circle, and if the points land mostly on the same half, the test is passed. This issue of non-specificity can be countered with using a high enough confidence level, and ensuring that there is enough data in the sample so that it has a decent chance of being uniformly distributed. Also, using the phase-sensitive version when possible helps. Running Rayleigh test in a moving window can help identify sections of the data where the processes are in sync (as in (Patel et al., 2009), see chapter 3.3.2), but given the low power of the test, and the large number of consecutive statistical tests that one ends up conducting, it is important to be careful in interpreting the results. Rather than using the Rayleigh test p-values from these windows as a threshold for synchronisation detection,

the whole trial should be considered (e.g. in a permutation test as in (Patel et al., 2009)).

Naturally, the upside of this unspecificity is that the test results hold regardless of the shape or type of the unimodal distribution the data is from (circular normal, cardioid, wrapped stable...), as it basically just rules out the uniform one. However, it is important to note that as the test is based on the \bar{R} score, it will only work on the unimodal distributions, and not on the bimodal or other multi-modal distributions. This is due to how \bar{R} is calculated, as two vectors facing opposite directions would cancel each other out, a bimodal distribution with identical peaks at 0 and 2π would also cause each other to cancel out, resulting in $\bar{R} = 0$, and the Rayleigh test giving a false negative. The same applies to all multimodal distributions.

5.2 Cross-recurrence analysis

5.2.1 Phase space and cross-recurrence

When studying music performance, the question is usually not whether the performers are synchronised, but how and to what degree. However, when studying, for instance, the responses of music listeners, or entrainment of postural sways, or some other index of interaction in the absence of clear, rhythmic and periodic framework, more sensitive measures are needed. In these cases, the assumption of periodicity might not be fulfilled.

Studying relative phase or cross-correlations between inter-tap intervals are linear analyses, in that they assume the interactions between the individuals are additive in nature. However, it can be argued that the interactions are non-linear and multiplicative (Shockley, Butwill, Zbilut, & Webber Jr, 2002). This would seem to be the case especially in absence of strong metrical schemata or other “anchors”, as may occur in free improvisation. Dynamical systems research has developed tools for analysing non-linear systems and even very weak forms of interaction and synchronisation. One such method is recurrence and cross-recurrence analyses. Recurrence is a fundamental characteristic of dynamical systems (Marwan et al., 2007), meaning that the systems tend to return to (or close to) an earlier state after a while. Recurrence plots were suggested as a tool for analysing patterns of recurrences in systems in the late 1980’s (Eckmann, Kamphorst, & Ruelle, 1987).

The evolution of a system can be visualised by plotting its phase trajec-

tory. According to Takens theorem, we can reconstruct the behaviour of a multidimensional system from just one univariate time series that we have observed (Takens, 1981). Thus, for example, the complex system that keeps our body in balance when we stand up can be reconstructed by observing the anterior-posterior movement of our chest. This is done by constructing embedded vectors from the time series and time delayed versions of it.

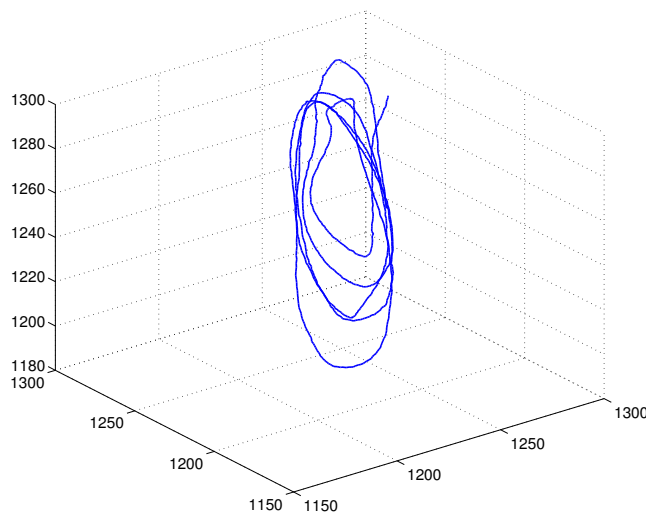


Figure 5.7: The phase space reconstruction of an African dance trial - 10 second sequence of vertical head movements, three embedding dimensions with a delay of 20 frames (0.2 seconds) are used.

The phase space in figure 5.7 seems to be approaching a limit cycle suggesting the movement is periodic. The periodicity brings about this circular form of the phase trajectory, as the process periodically revisits the same regions of the phase space. Recurrence plots visualise this characteristic of dynamic systems: in these plots, all those points are plotted where a phase trajectory comes close to the place where it has already been.

Figure 5.8 shows recurrence plots of three different processes. The time series data are plotted on top, and corresponding recurrence plots below it. The random process (right panel) produces a recurrence plot that looks like static on TV: as the time series is random, so are the recurrences. At the

other “extreme” is the sinusoidal data (left panel) that produces a clear grid. In both cases, however, the whole plot is homogenous and self-similar, which are indications of stationarity. (Marwan et al., 2007).

In the sine wave recurrence plot (left panel), the recurrences form equidistant lines. This is an effect of the clear, constant periodicity in the data. The distance of the lines corresponds to the period in the data. Also note that in the sine recurrence plot, all black dots, all recorded recurrences, occur as parts of larger diagonal structures. This is a reflection of the determinism of the function, and that basically at those lags the processes recur constantly from start to finish. (Marwan et al., 2007)

The tapping data here is a concatenation of 16 different tapping trials from one participant in one experiment. The inter-tap intervals from these trials were just laid end to end to get a time series that was of comparable length to the other two (middle panel). The horizontal and vertical lines seem to fall on boundaries of different trials, where larger tempo changes have happened (as the tempo had gradually drifted during the previous trial, and then was reset to standard in the beginning of the next one). They are indications of discontinuities in the data. There are also diagonal structures which are a reflection of temporal dependencies within the tapping data - drifts and periodicities in the variability around the mean ITI. (Marwan et al., 2007).

In cross-recurrence analysis, a pair of signals are investigated in a similar fashion: their phase trajectories are first reconstructed, and then noted where they jointly recur. As in any descriptive analysis, the plots can be very revealing, but interpreting them can be difficult and especially comparing plots to each other might be tricky. That is why methods for quantifying the structures of these plots were developed in the late 1990’s.

5.2.2 Cross-recurrence quantification (CRQ)

Cross-recurrence quantification analysis (CRQ) was developed by Zbilut, Giuliani and Webber (Zbilut, Giuliani, & Webber, 1998) and extended by Marwan et al. (Marwan, Thiel, & Nowaczyk, 2002). CRQ quantifies relevant features of cross-recurrence plots, such as in figure 5.8 so that they can be compared to each other. As was seen in the previous section and figure 5.8, different data produce very different looking recurrence plots. The amount of recurrences is not the only difference, however, but the way

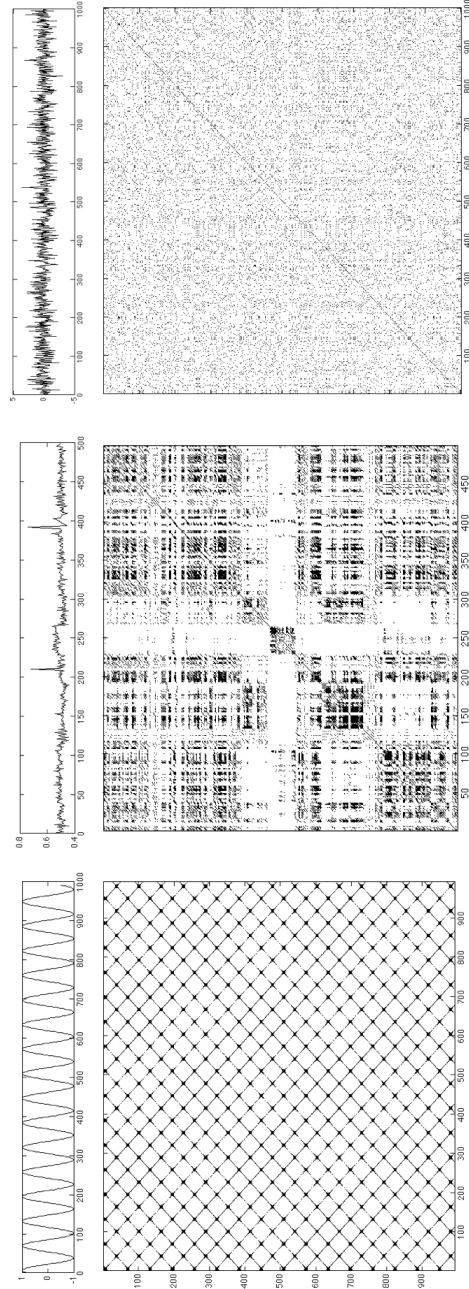


Figure 5.8: Recurrence plots of three different processes. The time series are plotted at the top, and recurrence plots of each of the processes in the large panels below. In the left panel, sine wave; panel in the middle: data from tapping trials; on the right, random sequence. (Figure inspired by Marwan et al. 2007)

the recurrences are organised (to vertical/horizontal or diagonal structures etc.) is significant. CRQ has measures for all these features, and they have been validated in many simulation and empirical trials, using non-linear oscillatory devices. They have even been applied in human interaction studies (Marwan et al., 2007; Shockley, 2005; Shockley et al., 2002; Webber, Marwan, Facchini, & Giuliani, 2009).

From a synchronisation detection and analysis point of view the most interesting measures are a) the overall percentage of recurrence, b) amount of determinism, c) length of longest diagonal line, and d) overall entropy. A full description of these measures and what they measure can be found in the excellent review by Marwan et al. (2007), so I will give only a brief introduction. The percentage of recurrence, or recurrence rate, is the ratio of recurring time points to the total duration of the analysed data, or in other words this is a density measure of the recurrence points in the recurrence plot. This can be used as a general synchronisation measure, as the more recurrence there is, the closer the two trajectories are to each other, indicating that they evolve in a similar fashion.

The amount of determinism measures the proportion of the recurrence points that are part of diagonal line structures. The more this happens the more one signal determines the values of the other, as their values co-vary. This can be seen especially in the case of sine wave recurrence plot in figure 5.8. The length of the longest diagonal line tells us what is the longest time these two signals have parallel trajectories in the phase space, and is thus a measure of stability of their relationship. To illustrate these measures, think of two trajectories that cross each other but are headed to different directions. This would be a chance cross-recurrence, so it would count towards the recurrence rate, but it would not tell that much about synchronisation of the two systems. So while it would add noise to the recurrence rate measure, it would probably not contribute to the amount of determinism or the longest line length, as it would be just a one-off chance recurrence. In contrast, if the two processes are synchronised, the trajectories will be parallel for a longer time. This will result in a longer row of recurrence points, oriented along the diagonal and thus adding to the determinism proportion as well as possibly towards the longest line. (Marwan et al., 2007).

Entropy of the diagonal lines is a measure of complexity in the recurrence plot. For example, entropy of the line lengths is very low in the random data

(all lines are short, one or two dots), which reflects low complexity of the underlying system (Marwan et al., 2007, 265). There are other measures of CRQ available, but they are not as central in synchronisation detection, and thus will not be discussed here.

Selecting parameters for the computation is a key step in the analysis. Methods for selecting these parameters are discussed in more detail in appendix D.

5.2.3 CRQ results and interpretation

As mentioned earlier, when analysing music or dance performances, we often know there is synchronisation, and know from the musical tactus its most salient periodicity. There are many cases, however, where the performance and the interaction is not stationary (or where it cannot be split into stationary segments for analysis). This might be the case in free improvisation, or in a music therapy session. In both cases the performance as well as the interaction can be non-stationary. While a therapy improvisation often requires a proper, hand-made musical analysis, automatising part of this process using data-driven methods (Luck et al., 2006, 2008) can be very useful in providing insights about the improvisation.

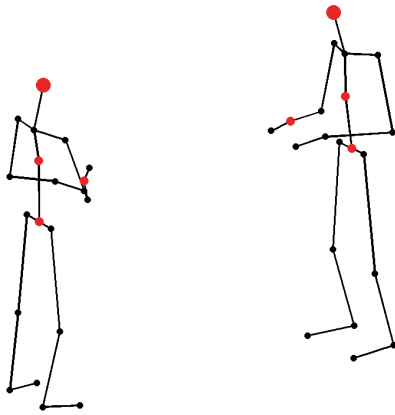


Figure 5.9: Joint markers analysed in the communication experiment. The ones coloured red were the ones focused on in the cross-recurrence analysis.

I will present an example of use of cross-recurrence quantification in musical interaction research. This example analysis comes from an ongoing

project where linguistic and musical interaction are compared⁴. The data are from two participants who are either telling a story together or improvising music together, using shakers. Their body movements were recorded using motion capture, and the interactions were also analysed qualitatively from the three-camera videos. The CRQ analysis is modelled after the body sway analysis in (Shockley et al., 2003), and also taking advantage of the example of parametre selection process and analysis in Shockley (2005). The example is from interaction where there is no clear periodicity in the movements, no beat-based entrainment can be observed, and there is no phase-locking in the participants' movements. For those situations, other methods, such as circular statistics are a better option. The strength of CRQ is that it can be applied to situations where other methods would be too "weak" to find anything.

The two participants are standing in a shared space. During the interactions, we expected the two participants' body movements to be entrained during moments of resolution or agreement. According to one of the theories behind the project (Gill, 2007), these "bouts" of synchrony can be very short. Consequently, we expected the movements to be strongly non-stationary, as the participants were exploring the range of expression of the shakers. We looked at the chest, root, head and hand markers, and used normalised position data. The examples presented here contain 20 second segments from what were originally two-minute trials. Rather than expecting there to be entrainment throughout these segments, or that the two participants would phase-lock their movement patterns, we used CRQ to characterise their phase relationships at greater detail during these segments, and then compared the "scores" of different segments to each other. Thus the length of the analysed segment is not as crucial as in the Rayleigh test, for example (see section 5.1.7 above).

After choosing appropriate parametres for the analysis (see appendix D for details), we calculated cross-recurrences from two trials, comparing the pair facing each other to not facing each other.

Figure 5.10 shows a comparison of the resulting cross-recurrence plots. The time-series are on top, and cross-recurrence plots below. Time advances

⁴The project is a collaboration with Dr Satinder Gill and Dr Marc R. Thompson. I carried out the cross-recurrence analysis in preparation for the 12th ICMPC & 8th ESCOM in Thessaloniki, Greece (Gill, Himberg, & Thompson, 2012).

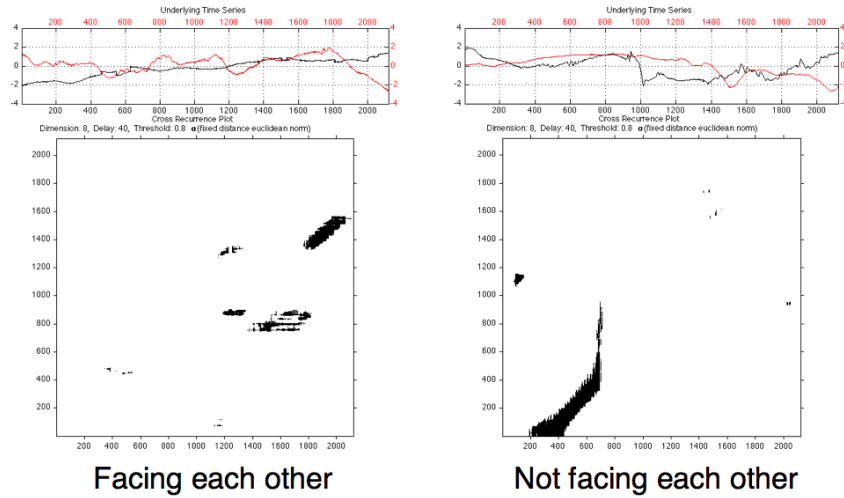


Figure 5.10: Cross-recurrence analysis example, using body sway data from a dyadic improvisation task. Panels show a contrast between a condition where the pair is facing each other and when they are facing away from each other.

along the diagonal from bottom left to top right. Black colour in the plot depicts cross-recurrence. Here, in the right panel, depicting a trial segment where the two participants are not facing each other, there is a long, diagonal structure at the beginning of the trial: here the two participants' movements were moving along a shared trajectory in the phase space. As the black, diagonal structure is not on the diagonal but below it, it means that one participant's phase was following the other with a lag. In the left panel, representing a trial segment where the two participants are facing each other, there is a shorter and less "fat" diagonal structure towards the end of the segment (top right corner), again indicating a bout of shared trajectory. These two examples represent an extreme - these segments are from the pair that had the largest difference in determinism across these two conditions. These are not the highest determinism or the lowest in the data, but rather, they show the largest within-pair contrast.

As can be seen from figure 5.10, most of the cross-recurrences are parts of larger structures, indicating that they are not just random recurrences but there is an actual connection that aligns the phase spaces. Determinism (proportion of recurrences in structures from total recurrence) was very high

in this data, in all trials it was at or over 90%, indicating that when cross-recurrence happens, it happens because one partner starts to “drive” the other, or they drive each other. Recurrence rate was between 1–5%, just as was hoped when setting the parameters of the calculations. Absolute levels of recurrence rate can be easily manipulated by parameter selection (see appendix D), but comparing them across trials can still be interesting; in this relative approach, recurrence rate reflects the amount of similarity or coherence between the two time-series.

Basic statistical analysis (t-tests) were conducted using the variables obtained from the cross-recurrence quantification. Facing and not-facing conditions were compared, and data from head, hand, chest and root markers were used. Although the cross-recurrence plots show interesting structural differences between trials, these changes do not seem to be strongly associated with the orientation and availability of the visual communication channel. The only statistically significant result was that the average diagonal line length was longer in the not facing -condition. In other words, to simplify, when they synced together, they synced for longer when they did not see each other. The data in figure 5.10 illustrates one example of this. This links very well together with the initial results of the qualitative analysis: people seem to be more comfortable and more likely to interact if they do not see each other. Standing face to face seemed to “freeze” some of the participants so that they were much more reserved in that condition, whereas they could interact musically and verbally more openly when they had the “privacy” offered by the condition where they did not need to look at each other. All information that was crucial for the task was conveyed via the auditory channel.

5.3 Summary: relative phase

In this chapter, I have presented and demonstrated the use of two methods for investigating phase relationships. The first approach, circular statistics, is a set of methods for analysing angular data. Using stroboscopic observation, discrete onset times can be converted to phase values, and their circular distribution can be measured and visualised. These methods provide a good way into looking at behaviour that is periodic, in relation to that periodicity.

The second approach, cross-recurrence analysis / cross-recurrence quan-

tification is somewhat difficult to use, as it requires a rather good familiarity with the dynamic systems literature and also quite a lot of preparation and plotting to find good parameters. On the other hand, it tolerates non-stationarity very well, and does not pose other stringent criteria for the data, either. Also, it is very sensitive for even fleeting and weak synchronisation, and synchronisation it detects does not need to be beat-based or periodical. While too cumbersome for analysing most rhythmic entrainment data, it is a useful set of tools to look at situations where entrainment is not as obvious. It expands entrainment analysis outside the periodic, rhythmic range of behaviours and into e.g. linguistic interaction, and unintentional interpersonal influence.

In the next chapter, I will discuss a mixed bunch of other methods that can be useful in analysing interaction and measuring influence. What these methods have in common is that they all can be used in switching the focus of the analysis from the *egocentric* measurement of individuals to *interactionist* analysis of the interaction between many participants.

Chapter 6

Analysing rhythmic interaction and influence

6.1 Introduction

In this chapter I will present and discuss methods for analysing interaction in dyads and groups, and estimating the amount and direction of influence in these interactions.

First, I will present pairwise intra-class correlation, which is a way to deal with “traditional” test-score data collected from dyads. Although this method differs from entrainment analysis, it can be useful for dealing with auxiliary data collected in entrainment studies, such as post-experiment ratings of liking or task difficulty. This method also helps to illustrate why it is not so simple to break away from the traditional methodology of testing individuals. In an interactionist approach (see chapter 2.6), the focus should be directed from the individuals to their interaction. However, in many cases it is not feasible to change what is being measured (e.g., it is often necessary to use standard questionnaires or measures), but at least it is possible to adjust how the data is being analysed.

Following this, I will discuss smoothing, which is a simple time-series technique that is akin to using the zoom in a camera, making phenomena at different time scales salient in time-series. This is followed by autocorrelation-based periodicity analysis, and cross-correlation. These are very common time-series analysis methods, and there is plenty of literature on both, so I will focus on how to use them in analysing rhythmic entrainment. The

last analysis method I present is Granger causality, a method often used in econometrics for measuring influence between two time-series.

This is not an exhaustive list. In neuroscience, a range of phase coherence and similarity measures are used to study synchronicity or similarity of activation patterns across interacting brains (Lindenberger et al., 2009; Nummenmaa et al., 2012, see e.g.). In music-related movement studies, estimates of mutual information have been used as an index of interaction (Braun et al., 2011; Papiotis, Marchini, & Maestre, 2013), and in dynamic systems research, a number of modelling and analysis methods have been developed for studying synchronisation in physical systems (Pikovsky et al., 2001). I have included in this chapter methods that I have used myself. At the end of the chapter, I will provide some examples of these methods in action.

6.2 Pairwise intra-class correlation

Pairwise intra-class correlation is a simple method for analysing social influence in the case where the outcome measures are for example test scores, numbers of interactions etc. rather than time series data such as tap times or movement data.

The intra-class correlation with visualisations and models for experiment design are nicely summarised by Griffin and Gonzales (Griffin & Gonzalez, 2003). They note that the individualistic approach of modern social psychology is at least partly due to methodological limitations. In this approach, individual behaviours are measured, and to maintain the level of control required by contemporary experimental psychology, any social interactions are controlled as well, for example in that apart from one participant all others in the social situation are confederates of the experimenter, or that the participant interacts with a computer program with preset parameters of interaction.

If interaction were to be studied *in vivo*, a few problems would emerge. In addition to the loss of control over many variables, the setup where many participants are tested at the same time can also lead to issues of statistical dependence of data collected from interacting participants. This lack of independence of measurements is usually considered a “statistical nuisance”, as the statistical dependence of the scores of the participants in the dyads or groups means that you do not meet the criteria for statistical tests on

the individual level. Instead, you will need to do the measurements at the dyad or group level, which means that you will lose a lot of statistical power with the degrees of freedom, as your n gets halved or down to a fraction from what it would be if you could measure performance at the individual level. Thus, Griffin and Gonzales have been promoting a way to quantify and visualise the within group correlation in relation to the between groups correlation, so that the contribution of the social influence can be assessed (Griffin & Gonzalez, 2003).

One way to investigate this interdependence is through interclass correlation (classic Pearson correlation), but this requires that the members of the dyads or groups can be somehow meaningfully labeled to independent classes. So, for example, if the dyads consist of a musician and a non-musician, or men and women, as in the example used by Griffin and Gonzales, it is possible to just correlate the scores of musicians and non-musicians, and see if they correlate. In this case, the two classes are presumed to have different means and variances, and correlations would suggest that the dyad membership has an effect on the scores.

But often the members of the dyads are not distinguishable this way - allocation of participants into dyads has been based on their musical training (thus members in the dyad are both either musicians or non-musicians) or is blind to their gender. In these cases, the pairwise intraclass correlation (PICC) is a useful method for getting a measure of social influence. The principle of PICC is very simple, but it can be connected to much more complex analyses of both intra- and interclass dependencies; here, I will only provide the simple example.

Imagine one has four dyads. For the PICC, we calculate the correlation using all possible within-group pairs of scores. As you see in table 6.1, we list the scores of our participants into one column, and in the reversed order in the other. Taking the Pearson product-moment correlation of these two columns will give us the pairwise intraclass correlation.

To interpret the correlation results, let us consider what the extreme values and zero score would mean:

- PICC will be 1, if all participants have an identical performance as their partner in the dyad, but the dyads differ from each other - all the variance is between dyads and none within.

Table 6.1: Example of pairwise intraclass correlation data handling

<i>Variable</i>		
	X	X'
Dyad 1	A1	A2
	A2	A1
Dyad 2	B1	B2
	B2	B1
Dyad 3	C1	C2
	C2	C1
Dyad 4	D1	D2
	D2	D1

- PICC will be 0, if there is as much variance between and within dyads: no evidence for similarity or dissimilarity across paired individuals.
- PICC is negative, if ratings vary more within dyads than between, groups are more dissimilar, and individuals within dyads are behaving in a complementary fashion.

A way of visualising this is to make an “antenna graph”, as in figure 6.1 by plotting the original data order against the flipped data order, and joining the scores of the pairs together, and drawing also the identity line (the diagonal). This gives an idea of both inter- and intragroup correlations. High intragroup correlations would be marked with pairs being close to the identity line.

Another type of visualisation that makes it possible to visually investigate the intraclass correlations is to plot the group scores on the y-axis with the groups ordered along the x-axis. This makes a plot that looks like a vertical abacus and intragroup correlation can be analysed by looking at how the “beads” or data points on each vertical column group together.

6.3 Smoothing

Maduell & Wing (2007, 616) show a fabricated example of how one person in a dyad would lead the other by initiating tempo changes in advance of the other. Plotting the inter-onset intervals is a good way of showing which participant initiated changes, be they intentional (expressive changes etc.) or unintentional (timing errors). Real data is unfortunately more messy than

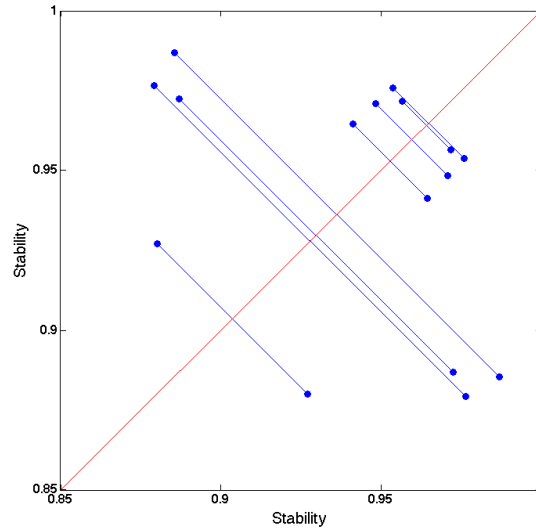


Figure 6.1: Example of a intragroup correlation visualisation with a “Antenna graph”. Here, stability scores from one cooperative tapping trial is plotted against itself, with the two participants’ scores flipped around. For highly intercorrelated data, the data points should fall on the diagonal axis (red line). Here, only data from a few trials do.

the imaginary one used in the example of Maduell & Wing. The top panel of figure 6.2 shows the inter-onset intervals of two tappers in a cooperative tapping trial. From this raw data it is very difficult to see anything specific about the relationship of the two series, due to the noise in the data. For example, while the tempo seems to be fluctuating during the trial (the average ITI seems to fluctuate), it is not easy to see who initiates these changes, as the noise, or the zig-zagging of the tap-to-tap variability obscures the larger trends.

We can improve the visibility of global changes by removing some of the noise, the local variability, by *smoothing* the time-series with a moving average. In this process, individual ITI-values are replaced with a moving average. The moving average is calculated over a segment, or a window, of the ITI’s. The size of this segment or window, as well as its shape or internal weighting, depends on the data and the purpose: the longer the section, the smoother the resulting plot. For example, smoothing with the

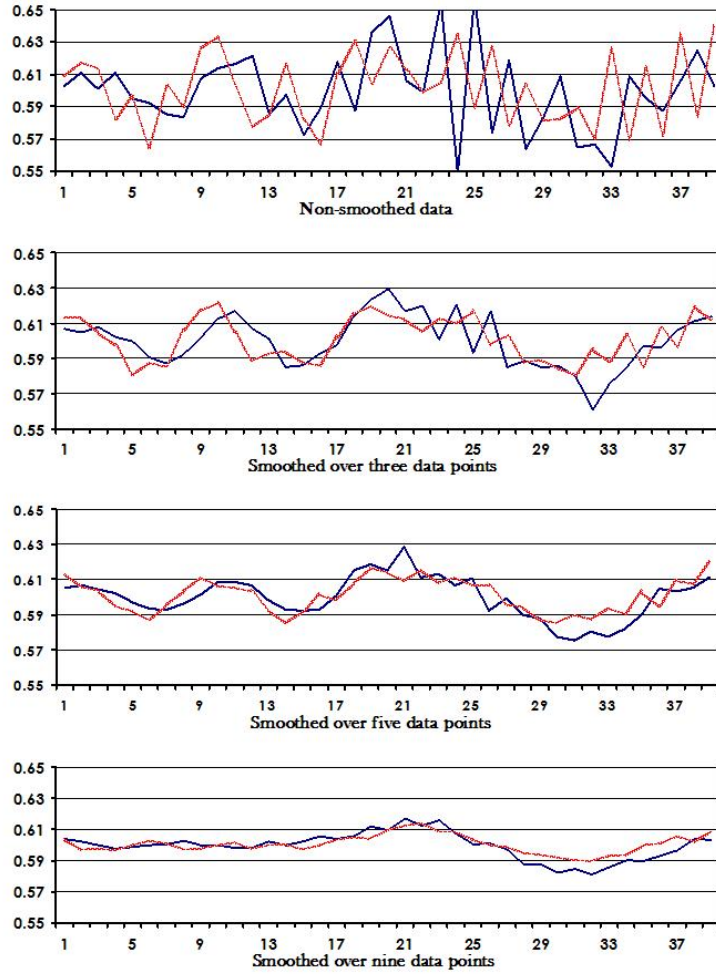


Figure 6.2: Examples of smoothing using different window sizes. From the top: not smoothed, window size 3, 5, and 9 data points (tap onsets)

moving average of 5 taps means that each ITI_n is replaced by the average ITI of $(ITI_n - 2 \dots ITI_n + 2)$. The averaging can also be weighted to emphasise the most recent values etc., but in this example, an un-weighted average was used.

In the second to fourth panels in the figure 6.2 the data has been smoothed using moving averages with window sizes of three, five and nine beats, re-

spectively. The bottom panel therefore presents data where each data point is an average of the tap itself and four taps before and four taps after it. Smoothing over three taps already starts to reveal how the interaction in this pair works; the red line of *Tapper 2* seems to be changing direction earlier than the blue line of *Tapper 1*, which seems to follow the changes one or two taps behind. For example, if we look at the speeding up and slowing down in the beginning of the trial (in the ITI graph, the shorter the ITI the faster the tempo, so when the curve goes up, tempo slows down, and vice versa), from ITI's 1–15, we can see that *Tapper 2* reaches the first valley (fastest tapping) at ITI number five, while *Tapper 1* gets there at ITI number seven, two taps later. The same happens again, at ITI:s 10 and 11 respectively, this time *Tapper 1* following only one ITI behind, when *Tapper 2* decides to stop slowing down and speeds up again.

This effect is clearest in the third panel, where the data is smoothed over five points. In addition to seeing the relationship between the tappers clearly, the curves maintain their initial shape, and it is possible to see the more global structure of the trial, the drifts and tempo changes. In the bottom graph, however, this is more difficult. After smoothing over nine points, the peaks start to erode and the valleys to fill up - and as a result it is also more difficult to pick up information about the dynamics between the two tappers. In this example, it seems that smoothing over five data points is optimal, providing an ideal combination of local detail and global structure. In different datasets, the optimal smoothing distance might be different; it depends on the variability and the global drifts in the data. Incidentally, the duration of the best window for this data, 5 beats, or about 3 seconds corresponds (very roughly) to the perceptual present or the upper limits of subjective metrization (London, 2012).

In this case, there was a clear reason for why *Tapper 2* was the one “in charge”. The trial was set up asymmetrically, so that *Tapper 2* heard the metronome throughout the trial, while *Tapper 1* only heard it in the beginning for eight clicks. Thus *Tapper 2* was in the position to tell whether the tempo is drifting away from the original tempo set by the metronome, and was able to make corrections, which *Tapper 1* then accepted and followed one or two taps behind. It is possible that the drifts were amplified by *Tapper 1* not hearing the metronome and thus not being able to be “helpful” to *Tapper 2* in maintaining a steady tempo.

6.4 Periodicity analysis

The data collected in any tapping experiment is initially time series data. This means that the observations are ordered in time. Financial information, such as stock valuation or interest rate, is a good example of time series data. In time-series data, the order in which these observations was made is important (Anderson, 1971; Chatfield, 2004; Shumway, 1988). With tapping, it may or may not matter. Often we are just interested in summarising the whole trial with one number (such as average inter-tap interval (ITI), or standard deviation of onset-time asynchronies) so that we can make comparisons across trial conditions. Sometimes, however, we might be interested in seeing how the ITI changes during the trial, as it can reveal how error correction works (Repp & Keller, 2008; Vorberg, 2005; Vorberg & Schulze, 2002; Vorberg & Wing, 1996). However, often in the error correction research it is assumed that the process does not change over time, and the trial is in fact a string of repeated measures. The experimenter is just interested in very local order of events, usually over two or three taps. However, with two tappers the interaction is usually not stationary. In other words, the relationship between the tappers changes over time, so that the means and standard deviation of any measure of this interaction would change as well. In these cases, using average scores might be misleading.

Autocorrelation analysis is often used to uncover periodicities in time-series data, for example, metre in music (Brown, 1993). It has been observed that in self-paced tapping, the adjacent ITI's are often different from each other, even when the average ITI is relatively stable (see e.g. Repp, 2005b; Wing & Kristofferson, 1973). This can be seen in the zig-zagging of ITI plots from most experiments (for example the raw data on the top of the figure 6.2).

This zig-zag it shows up as a negative lag one autocorrelation (adjacent ITI's are maximally different from each other). There can be positive correlations (self-similarity) as well. The bar-level grouping often shows as a positive autocorrelation at the lag that corresponds to the time signature¹.

¹Although positive autocorrelations do not necessarily reflect perceptually or behaviourally salient periodicities. For example, with periodic signals the “lower level” positive autocorrelations (for example at bar level) have “harmonics”, so that also the two-bar level a positive correlation can be found. This can be avoided by setting a maximum lag for the calculation, or by using enhanced autocorrelation that focuses on the first peak (Tolonen & Karjalainen, 2000)

For example, in the case of $\frac{4}{4}$ time signature, at the bar level, beats are grouped in groups of four, and the autocorrelation will show a positive lag 4 autocorrelation, whereas in $\frac{3}{4}$ time signature, there will be a positive autocorrelation at lag 3. Once the periodicities are known, the data can be further analysed using this information, for example using the periodicities in choosing referent periods in circular statistics (see previous chapter).

6.5 Cross-correlation

While smoothing gives us a way to visually inspect tapping trials and see some features of the relationship between participants, it is quite a crude method and does not show much detail, and the findings cannot easily be quantified. In contrast, cross-correlation of two time series can reveal dependencies between them, and provide quantitative results and estimates of statistical significance of the correlations (Boker, Rotondo, Xu, & King, 2002). When calculating the cross-correlation of two time-series, such as two series of inter-tap intervals, one calculates the correlation of the two, and then offsets one of the time-series by one data point, and calculates the correlation again, offsets it by another point etc. These offsets are called lags, and usually lags are calculated in both directions (left and right). Any significant correlations at non-zero lags suggest that one time-series is similar to the other one but with a time-delay that corresponds with the length of the lag.

In the case of two tappers tapping together, if one tapper is constantly following the other one with a delay of one tap (one tap behind), we can, by cross-correlating the inter-tap intervals of the two tappers, witness a significant cross-correlation at lag 1. Being one tap behind means that when one tapper starts to speed up or slow down, the other follows with a one tap delay; or if one plays a very early tap, the other reacts to it by being early on the next beat. This simplified concept of leading and following was discussed above in section 6.3.

While it is sometimes the case that one tapper dominates to such an extent that the “regular” cross-correlation yields significant results, it is much more common that the onus shifts from one tapper to another in a flexible and dynamic fashion. The mutual adaptation that occurs in interpersonal entrainment means that when we calculate the cross-correlation for a longer

section of tapping, for instance the whole trial, we often get correlations at both lags 1 and -1 , and so the cross-correlation alone can not answer the question “Who is leading?”, because the answer actually is “both” (Konvalinka et al., 2010). Konvalinka et al. noticed that in the conditions where only one tapper could hear the other, the one capable of following and reacting to the other showed lag 1 cross-correlation, and when both could hear each other, both had lag 1 cross-correlation, indicating they were following the other person, adapting to their tapping.

To reveal the fine structure behind this overall result, we can use the windowed version of cross-correlation. In this method, we calculate the cross-correlation in a window that sweeps across the time series from left to right. Thus, the cross-correlation is calculated in smaller sections, moving the section from beginning to the end of the time series in small steps. We can use different window sizes (the number of taps across which the cross-correlation is calculated each time), and different step sizes (the number of taps the window is shifted forward for the next round of calculations). Windowed cross-correlation can be used for discrete tapping data as well as quasi-continuous movement data.

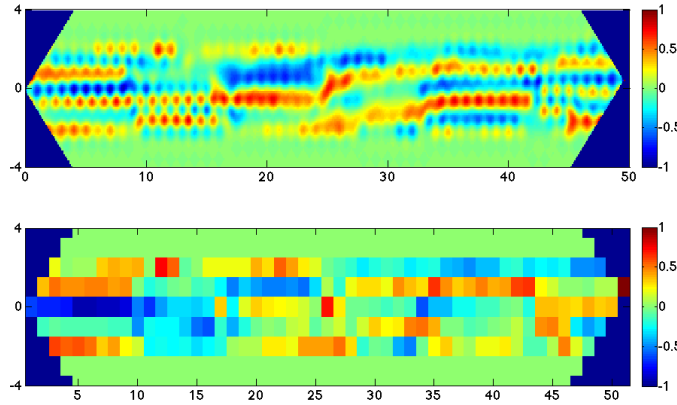


Figure 6.3: Example of windowed cross-correlations of a tapping trial. Bottom panel shows “Raw” correlation coefficients, the top panel has a “higher resolution” picture generated from the same data.

The result of these calculations is an array of correlation coefficients. There is a visualisation of one such array in figure 6.3. In the array, rows

represent lags while columns represent the iterations of taps, so that in the image, time flows along the horizontal axis, while the vertical axis represents the lags. The image is rendered and colours are used so that dark red indicates correlations close to $+1$ and deep blues correlations close to -1 . (The dark blue bands at the edges are an artefact of the windowing.) This figure can be read so that the data in the top half of the figure represent one tapper and the data in the bottom half the other tapper.

For instance, there seem to be sections where one tapper has followed the other at a lag of one tap, and in some other occasions the lag is larger. However, as there is both lag 1 and lag -1 activity, we can conclude that both participants were actually following the changes initiated by the other tapper, or in other words, the tappers were mutually adapting to each other. The top panel “high-res” picture is in this case a little bit misleading, but in many cases such visualisation is good for making sense of the structures of the trial.²

This demonstrates how error correction works through mutual adaptation in co-operative tapping. As Vorberg and Schultze (2002) and Repp and Keller (2008) have suggested, there is a co-operative mode of error correction, which requires both tappers to adapt their tapping in order to align them both to the mutually abstracted pulse. As every chamber musician or anyone who has played in a band, when one player misses a beat, comes in early or late, the whole ensemble needs to react, not only the one who made the error.

This mutual adaptation becomes second nature for musicians and is, just like phase error correction in general, mostly subconscious and automatic. It seems, though, that not everyone is doing it in the same way. In a study with autistic youths, we obtained co-operative tapping data, of which the windowed cross-correlation arrays are strikingly different from the ones drawn from the data of non-autistic participants. An example of this data is discussed below, in section 6.7.

²The cross-correlation function used to produce this figure in MATLAB has been written by Olivier Lartillot of University of Jyväskylä. It is superior to the one I wrote myself and which was used to produce the plots in figure 6.4.

6.6 Granger-causality

Granger causality (or G-causality) is a time series analysis method that has been used in economics and more recently in neuroscience (Granger, Ghysels, Swanson, & Watson, 2001; Seth, 2007). G-causality can be defined as follows: We have three terms, X , Y and W . Of these, X and Y are time series and W is a group of variables that we know have an effect on X . We then try to predict X_{t+1} with the previous values of X and the variables W , and then make another prediction using also the previous values of Y . If the latter prediction is more accurate, or in other words, if values of time series Y help us predict the future values of time series X , we can say that Y “Granger-causes” X . Applied to musical interaction, X and Y could be time series data obtained from a music therapy session, X representing the note onsets of the therapist and Y representing the note onsets of the client (we can ignore W for now). If we find that knowing X helps to predict Y (that X G-causes Y), then we can say that the client Y was influenced by the therapist, which might be of interest in the case of an autistic client.

The first step of the analysis is to estimate the model order, or the time lags at which X and Y are likely to influence each other. This can either be estimated based on *a priori* knowledge about the system, or by using either the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) (Seth, 2010, 264). Running the analysis then yields two test scores and associated p-values, one for X G-causing Y and another for Y G-causing X .

As mentioned above in the section on leading and following 4.3, in musical interaction, roles are rarely clear. Rather than having one leader influencing the other participant(s) constantly, the participants mutually adapt to each other, thus also mutually influencing each other. Granger causality helps to provide measurements of the amounts of influence that are exerted in each direction, thus providing a more graded measure of leading and following rather than just a categorical one. For example, in a music therapy improvisation, we can calculate the Granger causality both ways to see how the relationship between the therapist and the client is balanced.

There are a number of limitations for using G-causality. The most important one is that the time series should be stationary. Music performance data will not always meet this criterion. Formally, the criterion for stationar-

ity can be checked with a covariance stationarity check. And if the criterion is not met globally, the data can be analysed in smaller segments that satisfy this criterion locally. Thus, for instance, a windowing approach can be used, just as in the cross-correlation analysis. Also, normalisation or other transformation of data might help in preparing the data for this analysis; for example, in the case of movement, acceleration data might be more suitable than raw position data.

Other limitations include linearity; the data analysed should be linear, although extensions for non-linear data are being developed (Seth, 2007, 2010). Also, as usual when analysing correlation or causation between variables, the choice of variables and the validity of the model are key factors in the analysis: the mechanism of causation needs to be credible. This is analogous with the criteria for detecting synchronisation 2.2 (autonomy, interaction and coupling), where it was concluded that before engaging in statistical or quantitative analysis, it should be qualitatively explored if synchronisation is plausible and meets these criteria.

6.7 Examples - mutual influence

To illustrate how these methods for visualising and analysing mutual influence work, I am using a somewhat extreme example. This set of tapping data comes from a dyad where one participant is the experimenter ‘*C*’, and the other is a teenaged girl ‘*A*’ on the autism spectrum. The tapping data was collected in a student experiment that is discussed in more detail in appendix A.7. In the experiment, participants were asked to tap with a metronome, the experimenter or music.

‘*A*’ was one of two girls among the participants. She produced very steady tapping, but was not synchronised with either the metronome, music or the experimenter. She kept her initial, spontaneous tempo very accurately, even though it was much faster, and not metrically related to the beat of the stimuli. At this point we tested if *A* would mind if the experimenter synchronised with her. After she started another trial, the experimenter went along with her faster tempo and they performed a “continuation trial”, a duet tapping segment without a metronome.

The tempo (in average 173 BPM, ITI 0.346 s) was very steady, and the two participants were highly synchronised, as can be seen from the top

panel of figure 6.4. The ITI remain in a very narrow range all through the trial, apart from a couple of “mistakes” that the experimenter (‘C’, blue line) makes around taps 45–55. However, the participant (‘A’, green line) does not react to these perturbations but continues her steady tapping. This is surprising, as such phase attraction usually happens automatically.

The WCC plot (bottom panel) indicates that for most of the trial, the experimenter ‘C’ is the one who maintains synchrony by adjusting her tapping in response to ‘A’³. Most of the red colour, indicating positive correlation, is on the top side of the plot, corresponding to ‘C’ following ‘A’ at a lag of 1–4 taps. Almost no changes in A’s tapping seem to be reflections of what ‘C’ had done. This visualisation seems to confirm the observations of ‘C’ and myself that the ‘A’ was firmly sticking to her own beat and not synchronising with her partner.

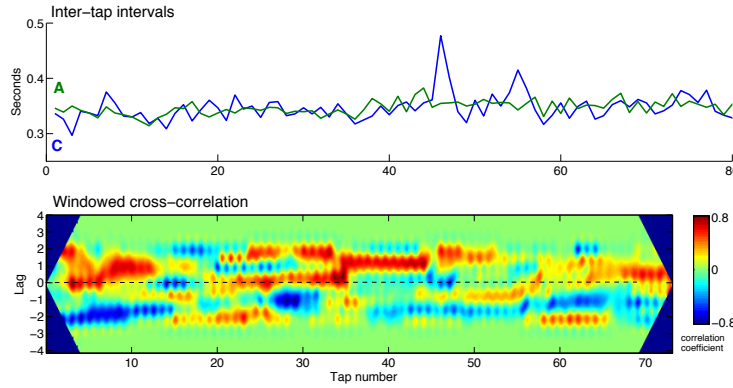


Figure 6.4: Example of a tapping trial, continuation tapping between A, an autistic girl and C, the experimenter. The top panel shows the inter-tap intervals, while the bottom panel shows the windowed cross-correlation.

This could be confirmed with Granger-causality analysis. First, the model order needs to be chosen. In this case, the windowed cross-correlation analysis can provide *a priori* information about the time lags of dependencies in these two time series. The figure 6.4 shows lags up to 4, and most high correlations occur at lags one through three. The AIC and BIC estimates are 3 and 2, respectively. Thus using the model order of 3 seems like

³Cross-correlation is calculated in a 10-tap window, with a step of 1 tap. This is why the WCC plot is 10 taps shorter than the ITI plot.

a rational choice.

Running the Granger-causality analysis for the inter-tap intervals of ‘A’ and ‘C’, using the model order of 3, we obtain two test results and associated p-values. For ‘C’s ITI’s Granger-causing ‘A’s ITI’s, $F = 0.2763, p = 0.84$. However, ‘A’s ITI’s G-causing ‘C’s ITI’s, we get $F = 8.5091, p = 0.0001$ confirming that ‘A’ influences ‘C’ but not vice versa. In this case the observation, visualisation (WCC) and the statistical test of Granger-causality all agreed - there was no mutual adaptation in the duet. What sounded and seemed like a highly entrained and synchronised performance, was in fact synchronous because just one of two participants was adaptive, and actively synchronising.

6.8 Group entrainment

6.8.1 Phase conversion: Hilbert transform

The concepts of period and phase are very general and applicable for all kinds of cases of periodic oscillation. Above, I have discussed how the relative phase of a discrete process can be estimated. This can be done for continuous data, as well. The instantaneous phase of a process can be estimated using a Hilbert transform (Khvedelidze, 2002), a method developed by David Hilbert in the early 1900’s. For example, we can take movement data, the vertical acceleration data from two participants dancing together, use the Hilbert transform to obtain the phase values for the movements and then simply subtract the phases from each other to see their relative phase. This can then be visualised using a rose histogram, and average θ as well as \bar{R} can be calculated.

The Hilbert transform was developed for periodic functions, and it assumes the data to be stationary and linear. Instantaneous frequency (and phase) can be obtained from nonstationary and nonlinear data using a so-called Hilbert-Huang transform, a method developed later by Huang et al. (1998). Both are implemented in the MoCap Toolbox for analysing movement data (Toivainen & Burger, 2010).

6.8.2 Kuramoto model

What if we have more than two entrainers? How could we analyse entrainment in a group of dancers, for example? One way of doing this is to look at each pair of dancers at a time and use the windowed cross-correlation or other tools discussed above. Another is to use a method that allows the whole group to be analysed as a whole, namely the Kuramoto model (Acebrón, Bonilla, Vicente, Ritort, & Spigler, 2005). The Kuramoto model describes synchronisation of a group of weakly coupled limit-cycle oscillators, and it can be used to measuring synchronisation in a group of musicians or dancers.

$$\frac{d\theta_i}{dt} = \omega_i + \sum_{j=1}^N K_{ij} \sin(\theta_j - \theta_i), \quad i = 1, \dots, N. \quad (6.1)$$

Equation 6.1 shows the dynamics of the oscillators θ_i in the Kuramoto model. Each oscillator in the group has its own natural frequency ω_i , and they try to run independently at that frequency. At the same time, the oscillators' coupling to others in the group (coupling matrix K_{ij}) tends to synchronise them together. The coupling strength between a pair of oscillators depends on the coupling matrix and also sinusoidally on their initial phase difference $\theta_j - \theta_i$ so that the closer they are to each other in phase, the stronger the tendency to shift towards each other. The coupling matrix can be used to govern which of the oscillators are coupled to which others. For example, if we model a large group of dancers, we could set the matrix so that each dancer is coupled with their nearest neighbours but not those further away across the dance floor.

The Kuramoto model can be used in modelling interactions in large groups, and in analyses of experiments of group interactions. Using the model, we can calculate a measure that characterises the synchronicity of the whole group. This *order parameter* is an index that can get values between 0 and 1, ranging from disorder to perfect synchrony. Conceptually, it is very similar to the \bar{R} index used as a synchronisation measure in dyads. The order parameter can also be calculated in a moving window, giving a continuous index of how well the group is together. There is an example of this in chapter 8

A continuous measure of the order parameter from a group of dancers would allow for example the study of entrainment of the group as a function

of the music being played. Perhaps the pulse clarity of the music would turn out to be a factor, and the group would be better entrained (higher order parameter) when music with a high pulse clarity was played? For such a study, movement of the dancers would need to be measured. This could be done relatively simply with just one optical marker attached to the dancer's heads, or with a device (either a smart phone or a wiimote) with an accelerometer in their pockets.

6.9 Summary - analysing interaction

In this and the previous chapter, I have presented a number of methods that can be used for analysing data from dyads or groups. As more and more dyadic and group experiments are carried out, new analysis methods and new applications of old methods are invented. Currently, there are no canonical methods for analysing entrainment, and thus analysing interaction and interpersonal influence is especially difficult.

Entrainment in dyads and groups is based on mutual adaptation. This adaptation is dynamic in that as the interaction unfolds, the roles of leader and follower change constantly. Thus analysis methods that can show or quantify this dynamism are important; a cross-correlation analysis of a co-operative tapping trial that yields positive correlations at both lags 1 and -1, suggests that both participants are followers in the interaction. Conducting this analysis in a moving window reveals the alternating pattern of following, and the dynamic nature of this relationship.

Studying the interaction rather than just the mechanism of synchronisation poses challenges not only for the analysis methods but for the whole research cycle from the technical setup to the experimental procedure to the theories, hypotheses and interpretations of data.

In the next two chapters, I will present data from two empirical experiments. The first is a dyadic tapping experiment, comparing tapping with a human partner to tapping with non-responsive computer partners. This study introduced social dimension into rhythm and timing studies. The second study continued in expanding the horizons by including embodiment. The African dance study was an exploration of group musicking and dancing, using e.g. cross-correlation analyses to explore patterns of similarity in timings of dance movements.

Part III

Empirical studies and discussion

Chapter 7

Human vs. computer tapping partners

7.1 Introduction

As described above in chapter 2, human synchronisation and time-keeping abilities have been studied using a tapping paradigm for more than 120 years. In these studies, the accuracy in synchronising with an external referent or in keeping the original tempo after this referent has faded out is explored in a wide variety of conditions. The length of trials ranges from a few taps to many minutes, the micro-timing of the pacing signal (the referent) is often manipulated (Stephan et al., 2002), the feedback that the participants hear of their own tapping has been delayed or muted (Pfordresher & Palmer, 2006), and of course tappers of various ages and levels of musical training have been studied (McAuley, Jones, Holub, Johnston, & Miller, 2006).

In these studies, the research question is often a variant of the following: how do we keep time, how does our synchronisation mechanism work? There have been two main schools of thought about this; the cognitively oriented group has been looking for the mechanisms of an internal timekeeper and evidence for its functions as a part of the cognitive system (Mates, 1994; Schulze & Vorberg, 2002; Wing & Kristofferson, 1973), while the other group, taking influences from dynamic systems and neural networks, view synchronisation and entrainment as emergent phenomena that can be modelled using e.g. oscillator models (Large, 2000). More recently, improved imaging methods in neuroscience, and the use of motion capture

have spurred a new wave of research on synchronisation and entrainment in dyads and groups, this time based on theories of brain mapping, cognitive neuroscience, and movement science (Grahn & Brett, 2007; Luck & Nte, 2008; Toiviainen et al., 2010).

In almost all tapping studies to date, an individual participant is performing tasks paced by a metronome or a computer (the few exceptions are discussed in chapter 6). While this makes sense methodologically, it also means that only a narrow band of the full spectrum of synchronisation abilities has been studied. Most importantly, the social context is missing from these studies. The social context can be included by studying entrainment and interactions of dyads instead of experimenting with isolated individuals. The motivation for this has been discussed in chapter 2.5, and the tools for taking this step have been presented in chapter 5.

In the current study, the objective is to establish whether mutual adaptation has an effect on synchronisation accuracy and stability, and to investigate in general whether there is a difference between synchronisation to a non-responsive referent vs. an actual, interpersonal entrainment. An experiment, where participants were paired together, tapping with each other and with various computer tappers, was conducted. Participants had only auditory contact with each other (i.e. they could not see each other while performing the tasks, although they met each other before and after the tasks), and were not told beforehand whether the partner they tapped with at any given trial was the other participant or one of the computer tappers. The computer tappers were non-responsive, so they represented the traditional setting of synchronisation trials.

7.2 Methods

7.2.1 Participants

Data was collected in two separate sessions, first at the *Cambridge Centre for Music & Science* in August 2006, when seven pairs of tappers took part. A second, additional data collection was organised in Jyväskylä, in May 2009. In this second experiment, where the procedure and the stimulus were very slightly changed, nine more pairs were recorded.

7.2.2 Tasks and stimuli

Participants performed a series of short tapping trials together. All trials followed the outline of canonical synchronisation-continuation experiments, meaning that they all started with a pacing metronome that faded out after 8 or 12 taps, after which the participants continued with each other for about 15 seconds, until they heard a prompt to stop.

There were three kinds of tapping tasks: synchronisation, alternation, and rhythm tapping. In *synchronisation* tasks, the participants were both tapping “on the beat”; in *alternation* tasks one participant was tapping on the beat, while the other was tapping in between the beats; in the *rhythm* tasks the participants were tapping a simple, interlocking rhythm. As in the synchronisation tasks, both participants had the same tapping tempo (equal event rate) in the alternation tasks, but they attempted to maintain a 180 degree phase difference (rather than 0 degrees in the synchronisation task). The alternation trials were run twice, so that both participants would in turn get to tap on the beat as well as in between the beats.

The rhythm pattern was a simple, repeating $\frac{4}{4}$ pattern: participant one was tapping a rhythm where the second quarter note was subdivided to eighth notes $|X \ xx \ X \ X|$ while participant two had the third quarter in each $\frac{4}{4}$ bar divided into two $|X \ X \ xx \ X|$. Together these formed an interlocking pattern $|X \ xx \ xx \ X|$. Again, the tasks were repeated so that both participants got to play both parts.

An isochronous MIDI tone sequence with 0.2 second note durations was used as the pacing metronome. The timbre of the metronome was crafted in the synthesiser, using a high-pitched piano sound stripped of reverb. The rise and decay times of the sample were set to minimum, which resulted in a dry, percussive sound that would not necessarily be identified as a piano, but it would be easily distinguished from the sounds of the participants’ drums. An arpeggiated MIDI piano chord was used as a jingle that signalled the end of the trial.

In the first run in Cambridge, the metronome ran for 8 beats, and 12 beats in the second run in Jyväskylä. The tempo in the Cambridge run was 120 BPM, while it was 100BPM in the Jyväskylä run. Both tempi are within the range that most people find comfortable.

Participants would always hear a feedback sound from tapping their own drum. Reverbless woodblock sound were used. To allow the participants to

Table 7.1: Summary of the characteristics of the four different computer tapping partners.

<i>Partner</i>	<i>Phase variability</i>	<i>Period variability</i>
Deadpan	no	no
Phase variant	yes, similar to human tapper	no
Fast	Yes, similar as in phase variant	constantly speeding up
Slow	Yes, similar as in phase variant	constantly slowing down

tell their two sounds apart, one would have a sound from MIDI instrument named “Woodblock, high”, while the other had sounds from the MIDI instrument “Woodblock, low”. The computer tappers’ sounds were matched to the participants’ sounds, so that the computer tappers had the same high or low woodblock sound as the human tapper they “replaced”.

There were four different kinds of computer tappers (summarised in table 7.1). The deadpan partner was perfectly isochronous, metronome-like tapper, with constant inter-onset intervals. The phase variant partner had a steady tempo, but jitter was added to the timings of onsets, so that there was variability in its phase, comparable to high quality human performance (mean IOI = 0.6s, SD = 0.02s). These phase variant computer partners were constructed from recorded synchronisation-continuation performances by two musicians. These tapping trials were first averaged and then the phase errors amplified so that the phase variability in each individual example was similar to the average variability in the whole set, while any individual “outlier” taps were cleaned up. The resulting tracks have the usual negative lag 1 -autocorrelation structure of self-paced tapping (Wing & Kristofferson, 1973). In addition to the deadpan and phase variant tracks, two different period variant tracks were prepared. These were constructed from the phase variant tracks by multiplying the onset time sequence with a constant to achieve a 10% constant acceleration or deceleration of tempo by the end of the sequence. These “fast” and “slow” computer tappers thus had variability in both their phase and period.

In summary, the participants performed in seven different conditions:

1. With another human in actual interaction (both heard each other, mutual adaptation was possible). This condition hereinafter referred to as “Human”.

2. With a deadpan computer tapper (computer tapper had no phase nor period variability). “Deadpan”.
3. With a computer tapper with human-like phase jitter (phase variability but constant period). “Phasevar”.
4. With a computer tapper that was speeding up (phase variability plus constantly shortening period). “Fast”.
5. With a computer tapper that was slowing down (phase variability plus constantly lengthening period). “Slow”.
6. With a non-responsive human tapper following a computer with steady tempo (the other participant heard either the *Deadpan* or *Phasevar* computer tapper, thus was not responding to the participant). “Non-responsive steady”.
7. With a non-responsive human tapper following a computer with varying tempo (the other participant heard either the *Fast* or *Slow* computer tapper, thus was not responding to the participant). “Non-responsive periodvar”.

7.2.3 Apparatus

Cambridge, Aug 2006: Stimuli was presented and data collected with Digital Performer 5.1 sequencer software that was running on a Mac G5 PowerPC with OS X. The stimulus presentation was mixed from four different MIDI channels. Regardless of the condition, the participants would always hear the metronome + jingle -channel and the channel with the feedback of their own tapping. Depending on the condition, they would also hear either their partner’s tapping or a computer tapper. Roland XV-5050 sound modules and a Roland JV-50 synthesiser were used as digital instruments, and the audio outputs of the different channels were mixed to two streams using a Yamaha O2R digital mixing console.

The participants received the signal via Behringer Powerplay Pro 8 headphone amplifier that allowed the loudness of the individual input to be controlled without interfering with the balance of the individual channels within the mix. The participants had Beyerdynamic DT770 over-the-ear

headphones and were using Roland SPD-6 MIDI drum pads. The MIDI signal was routed to the computer using MOTU MIDI Timepiece USB-MIDI interface.

Jyväskylä, May 2009: Stimuli was presented and data collected with Logic Pro 8 sequencer software that was running on an Intel iMac with OS X. Logic Pro's native digital instruments were used for sounds, and mixing the two outputs for the participants was done in Logic's own software mixer/output control, and relayed to the participants via a Fireface 400 external sound card. The two output channels were then patched through a Soundcraft Compact4 -mixer that allowed the outgoing volume of each channel to be controlled easily without influencing the balance of individual tracks within the mix. The participants had Sennheiser HD 25-1 II headphones and were tapping on Roland Handsonic10 MIDI-drums. the MIDI signal was routed to the computer using an Edirol USB-MIDI interface.

The same computer tappers were used in both sessions. Although different synthesisers were used in the two experiments, as similar instruments were selected as possible for participant tapping sounds, and in both runs they were optimised to have a short rise time and no reverb.

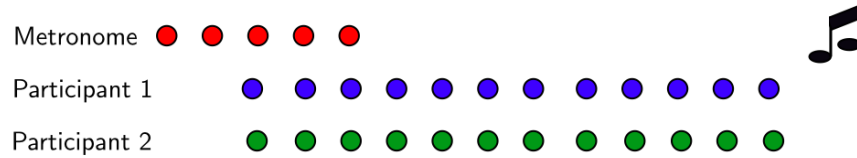
7.2.4 Procedure

Upon arrival, the participants filled in a one page questionnaire about their musical expertise and were given a brief explanation about the purpose of the experiment, tasks and the procedure. Participants were briefed together in the same room and their interaction was neither limited nor especially encouraged before the trial.

The experiment started with a short training session where the participants could familiarise themselves with the drums, sounds, and tasks. The participants were instructed to tap with one finger using their preferred hand, and aim for a sticker in the center of one of two pads in the MIDI drum (while all of the surface area of the pads is touch sensitive, the centers give the cleanest signal). The metronome and the end jingle were then presented and the volume of their headphones was adjusted to comfortable hearing level. The participants then practiced the synchronisation trial alone, and then together so that they both heard the metronome, themselves and each other's tapping (i.e. a simulation of the Human condition).

Each trial started with a metronome, and ended with a jingle. All trials

No computer tapper (“Human” condition)



Computer tapper included (all other conditions)

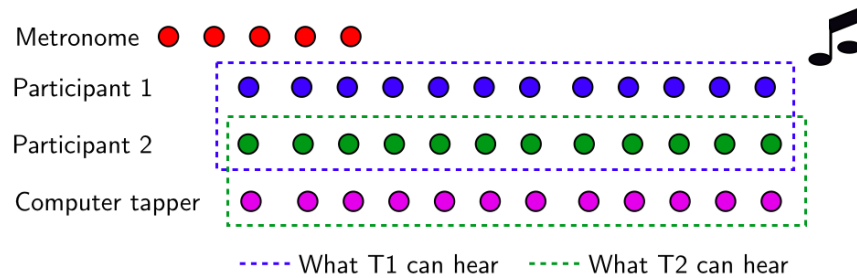


Figure 7.1: Structure of the tapping trials. In top panel, the human-human condition, where participants mutually heard each other. In bottom panel, the other conditions, where only one of the participants heard the other human participant, who in turn heard one of the computer tappers.

had the same tempo. Participants were instructed to start tapping as soon as they felt they got the beat. They were encouraged to keep tapping until the end jingle, regardless of possible mistakes or difficulties. The participants were informed that they would be occasionally tapping with computer tappers instead of the other participant, but they were not told when. In order to avoid the participants feeling frustrated, they were also warned that some of the computer tappers might be difficult to tap with.

Participants were instructed to try to keep the original tempo as well as possible and be as accurate as possible, and that they should also keep together with their tapping partner.

In all trials, both participants heard the metronome and the end prompt, as well as feedback of their own tapping. In the human-human condition, both participants could hear each other (see top panel of figure 7.1, condition number 1 in the list in section 7.2.2), and therefore be bidirectionally coupled to each other. In the other conditions, one of the participants (bot-

tom panel, participant 2) would hear one of the non-responsive computer tappers, and thus be unidirectionally coupled to it (conditions 2–5 in the list in section 7.2.2). At the same time, the other participant (participant 1 in the figure) would hear participant 2, and be unidirectionally coupled to him/her (conditions 6–7 in section 7.2.2, depending on what computer tapper participant 2 was following).

7.2.5 Analysis methods

Only the continuation section of the trial (the section between the fade-out of the metronome and the end jingle) was used in the analysis. Synchronisation and tempo drift in the trials were measured, using circular statistics. In both, concentration of the circular distribution (\bar{R}) is used, and the two measures differ only in the way onset times are converted into phase angles (as discussed in the methods section in chapter 5.1.5). I will subsequently refer to these measures as \bar{R}_s for the synchronisation measure and \bar{R}_t for the tempo stability measure.

We are mainly interested in the analysing how well participants synchronise with each other and with the computer tappers. In the synchronisation measure, participant’s phase values were calculated in reference with their partner’s (othe human or computer) onsets. The \bar{R}_s , concentration of this phase distribution, reflects the level of synchrony in the pair. This measure is insensitive to either the global tempo drift or the local variability of the inter-onset intervals, and only measures the relative phase of the two partners.

To measure how well the original tempo was maintained, the onset times of participants were converted to phase angles using the projected continuation of the metronome onsets as a referent (as in 5.1.5). This measure is sensitive to global tempo drift, as high \bar{R}_t values are associated with maintaining a steady relationship with the beat that the metronome initiated (high tempo stability), while any shifts in tempo, even if maintaining that new tempo very well, will result in a phase error that grows constantly, resulting to a very low \bar{R}_t . To keep the period variable conditions comparable with the other conditions, the referent in these cases was the projected continuation taking into account the 10% tempo change that the computer tapper was initiating.

Repeated measures ANOVAs were conducted for both measures, with

Table 7.2: The factorial design for the experiment (part)

<i>Partner</i>	Human				Deadpan				Phasevar				...
<i>Task</i>	S	A-on	A-off	R	S	A-on	A-off	R	S	A-on	A-off	R	...

the participant group as a between-subjects factor. The ANOVA model was thus 4 (Task) x 7 (Partner) x 2 (Group) full factorial with type III sums of squares. The four levels of the factor “Task” were *Synchronisation*, *Alternation on-beat*, *Alternation off-beat* and *Rhythm*. The seven levels of the factor “Partner” are summed in the table 7.1: *Human*, *Deadpan*, *Phasevar*, *Fast*, *Slow*, *Non-responsive steady* and *Non-responsive periodvar*. The trials with non-responsive human partners were grouped into two categories, depending on whether the other participant was tracking a computer that had a steady tempo, or one that changed its tempo. Part of the factorial design is depicted in table 7.2. The Greenhouse-Geisser method was used to adjust the degrees of freedom to counter for possible violations of the sphericity assumption of ANOVA, and Bonferroni corrections were applied to adjust for multiple comparisons in the pairwise comparisons analysis.

7.3 Hypotheses

The synchronisation measure \bar{R}_s and the tempo stability measure \bar{R}_t) were used as dependent variables in the statistical analysis. Based on earlier studies and pilot experiments, predictions can be made about how the task conditions affect the scores.

- c) \bar{R}_s **synchronisation** In synchronisation, predictability of the partner is important, so high values in the *Deadpan* condition are expected, but the actual interaction in the *Human* condition should also show up on top, especially when compared with *Phasevar*, given that in the former there are two people collaboratively correcting any timing errors (see Repp & Keller, 2008) whereas in the *Phasevar* the human participant has to do all of the correcting alone. The *Fast* and *Slow* conditions are probably producing unexpected errors more than the *Phasevar* condition, meaning that they are going to perform worse in this metric.

a) \bar{R}_t **tempo stability** As this measure penalises for even slight tempo drifts, tempo stability should be best in those trials where a participant performs with the computer tapper that keeps the correct tempo (*Deadpan* and *Phasevar*). As the participants were instructed to *both* maintain the original tempo *and* keep together with their tapping partner, in the *Human* trials \bar{R}_t might indicate which one of the two instructions was given the priority: lower scores would indicate prioritising keeping together over maintaining the tempo. In the *Fast* and *Slow* conditions, the referent takes into account the acceleration or deceleration of the computer tapper, so here high values indicate that the tempo of the participant is easily affected.

In summary, the expectation is that in synchronisation and tempo stability measures, the perfectly predictable partner, *Deadpan*, should yield the highest \bar{R} scores. Comparison between *Phasevar* and *Human* is interesting; the success of *Human* condition depends on how important the factor of mutual adaptation is in relation to pure predictability in these simple tasks. If mutual adaptation has no effect, then there should be no difference between the *Human* and *Phasevar* conditions.

7.4 Results

7.4.1 Synchronisation

For the synchronisation measure \bar{R}_s , the phase errors were calculated in relation to the partner, either human or computer.

Plots of figure 7.2 reveal very neat clusters in the *Human* trials, and that the unsynchronised trials in the middle are mostly from *Phasevar* and *Deadpan* conditions. *Phasevar* is somewhat more messy, while *Fast* and *Slow* ones are the most difficult, as can be seen from the lack of clear clusters. There are a lot of good values also in the non-responsive data, which suggests that it is not necessarily any more difficult to tap along a human tapping with someone else than it is to tap along with a playback or a computer tapper - none of them will respond or react, but humans are good synchronisers even in sub-optimal and unilateral conditions.

Significant main effect of *Partner* was found $F_{(5.4,51.6)} = 23.637, p < 0.0005$. The main effect of *Task* was also statistically significant $F_{(2.2,39.5)}$

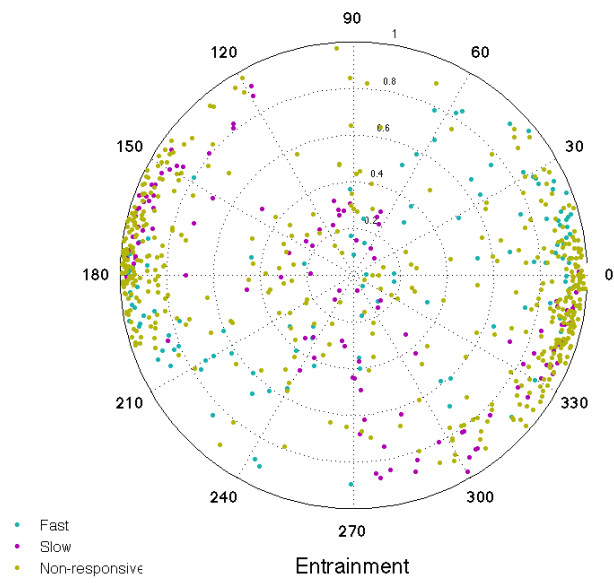
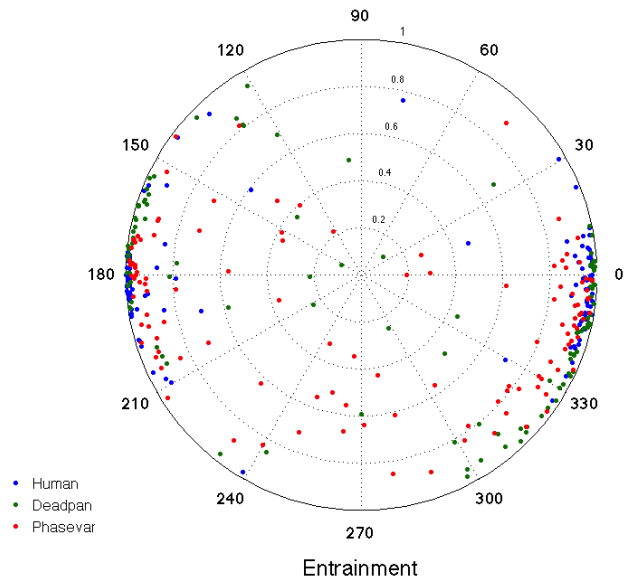


Figure 7.2: Polar scatterplot of the synchronisation measure. *Human*, *Deadpan* and *Phasevar* on the top panel; *Fast*, *Slow* and *Non-responsive* partners on the bottom panel.

$= 3.849, p = 0.026$, as was the interaction of the two main effects $F_{(7.3, 131.7)} = 2.863, p = 0.007$. Effect size estimates are $\eta^2 = 0.568$ for *Partner*, $\eta^2 = 0.176$ for *Task*, and $\eta^2 = 0.137$ for their interaction. There is also a statistically significant effect of *Group*, the between-subjects variable ($F_{(1, 18)} = 13.958, p = 0.002, \eta^2 = 0.437$). The *Group* also interacts with *Partner* (see figure 7.3), and there is even a three-way interaction between *Partner*, *Task* and *Group*.

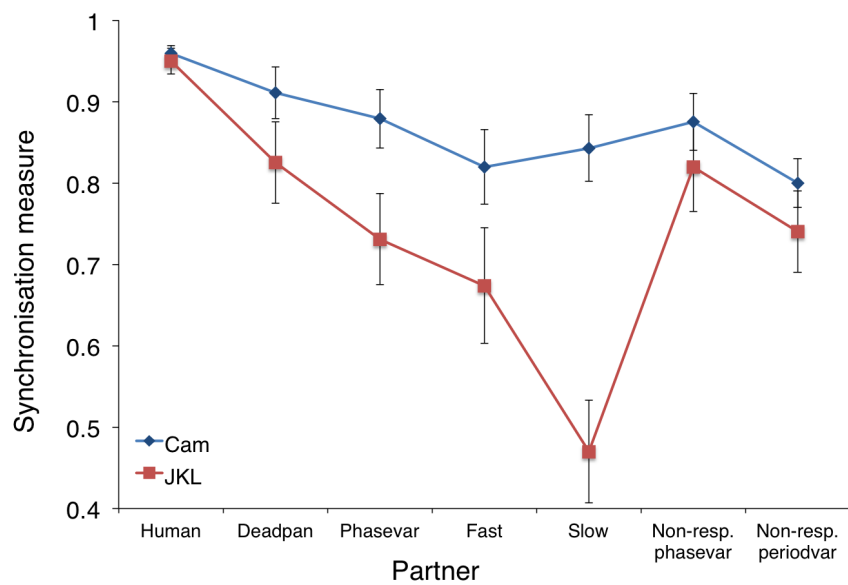


Figure 7.3: Synchronisation: Effect of *Partner* in the two *Groups*. Error bars represent standard error of the mean. Note the range of the y-axis from 0.4 to 1.

Looking at the *Partner* factor (figure 7.3), *Human* condition, the real interaction takes the first spot, and here there is no difference between participants in the two *Groups*, Cambridge and Jyväskylä. The Jyväskylä group scores lower synchronisation marks in all other conditions, especially with the *Slow* partner. This difference could be due to the slower baseline tempo for this group. Slower tempo leads to more variability (see e.g. Mates, Müller, Radil, & Pöppel, 1994), and this could be most evident in the condition where tempo is slowed down from the baseline.

In *Task* (see figure 7.4), the on-beat alternation is the most difficult task,

especially for the Cambridge group, who excel in the Rhythm task. While the Jyväskylä group performs more evenly across different tasks, the *Task* * *Group* interaction is not statistically significant.

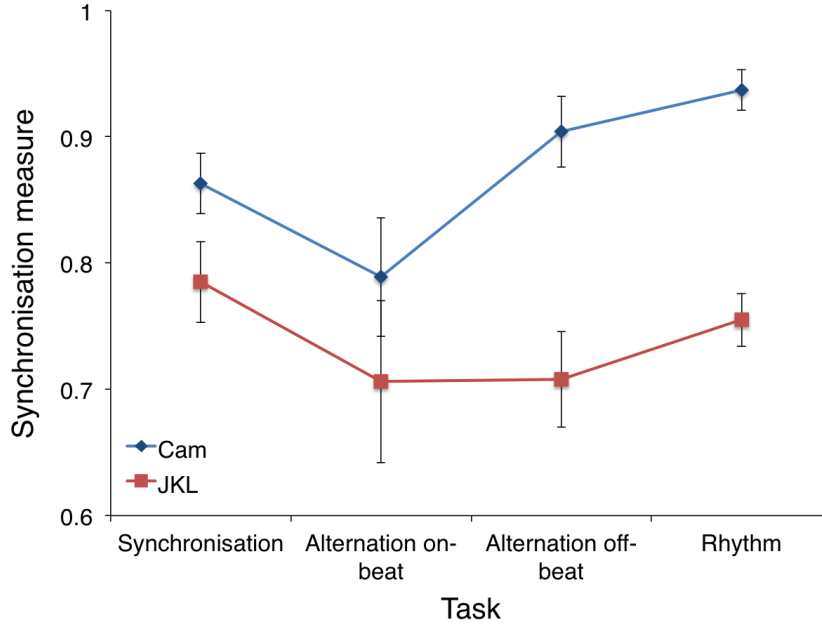


Figure 7.4: Synchronisation: Effect of *Task* in the two *Groups*. Error bars represent standard error of the mean. Note the range of the y-axis from 0.6 to 1.

Looking at pairwise contrasts (with data from both groups pooled together), the difference between *Human* and *Deadpan* conditions is not statistically significant (pairwise comparison $p = 0.151$). The difference from the *Human* to the *Phasevar* is statistically very significant ($p = 0.003$), however, as are the differences between *Human* and all the rest of the partners. In terms of *Task*, none of the pooled pairwise comparisons are significant.

7.4.2 Tempo drift

The highly synchronised *Human* trials (Figure 7.5, top panel) seem to be scattered around the circle without forming clear clusters, apart from around the centre, with relatively low values of \bar{R}_t . The human pairs were poor in maintaining the original tempo, compared to when the participants were

tapping with the *Deadpan* and *Phasevar* computer tappers. These data are clustered around the circumference, at 0 and 180 degrees (θ_t values of the *Alternation off-beat* trials are expected to be approximately 180 degrees). Also the *Non-responsive steady* trials cluster around the edges of the circle, demonstrating how helpful the computer tapper was in keeping the tempo. The *Fast* and *Slow* trials populate the centre of the circle as do the *Human* trials and *Non-responsive periodvar* -ones.

The repeated measures ANOVA showed significant main effects of *Task* $F_{(2.4,47.9)} = 5.300, p = 0.006, \eta^2 = 0.209$, *Partner* $F_{(3.8,76.7)} = 65.153, p < 0.0005, \eta^2 = 0.765$, and their interaction $F_{(7.0,140.4)} = 10.216, p = 0.007, \eta^2 = 0.338$. Also the interaction between *Group* and *Partner* was statistically significant $F_{(3.837)} = 8.359, p < 0.0005, \eta^2 = 0.295$. The three-way interaction between *Task*, *Partner* and *Group* was also significant, but the between-subjects effect of *Group* was not statistically significant.

Of the partners (see figure 7.7), as expected the *Deadpan* and *Phase* variant helped to keep the tempo best. The data suggests that even “second hand” influence helps in keeping the tempo, as the *Non-responsive steady* condition came in third. Again the Jyväskylä group is the more variable one, apart from in the *Human* trial condition. It is surprising that the actual interaction leads to as low tempo stability as with the period-variant partners or even the *Non-responsive periodvar* partner.

The pairwise comparisons of the pooled data indicate that the *Human* trials were statistically significantly more “drifty” than the *Deadpan*, *Phasevar* and the *Non-responsive steady* ($p < 0.0005$) but not different from the other conditions.

In terms of *Tasks* (see figure 7.6), tempo stability was highest in *Synchronisation* tasks and lowest in *Rhythm* tasks, especially in the Cambridge group. The *Rhythm* tasks with *Partners* that were changing tempo were especially unstable, which accounts for the statistically significant interaction of the two factors. Pairwise comparisons with pooled data suggest that the difference between *Synchronisation* and *Rhythm* was statistically significantly different ($p < 0.0005$), as was the difference between *Alternation off-beat* and *Rhythm* ($p < 0.02$).

Looking at tempo stability we found, unsurprisingly, that in those trials where the computer tappers were keeping the tempo steady, the participants also adhered to it, even when they needed to do so “second hand”, following

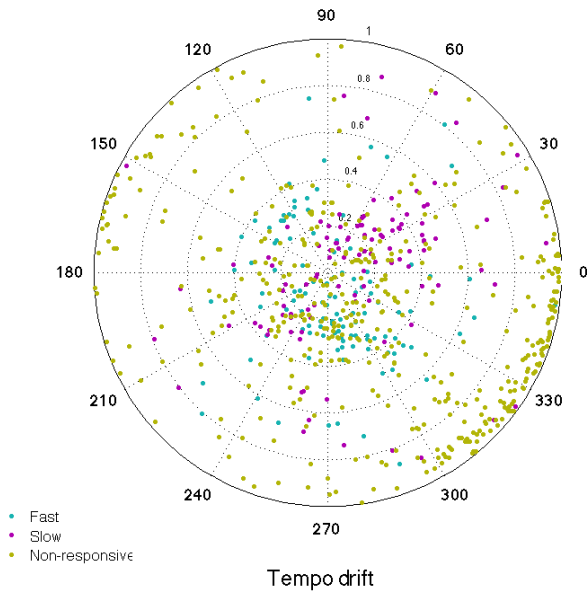
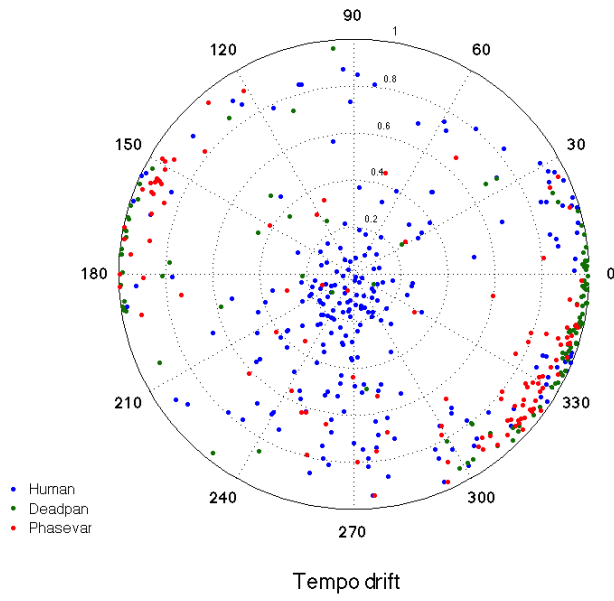


Figure 7.5: Polar scatterplot of the tempo drift measure. *Human*, *Deadpan* and *Phasevar* on the top panel; *Fast*, *Slow* and *Non-responsive partners* on the bottom panel.

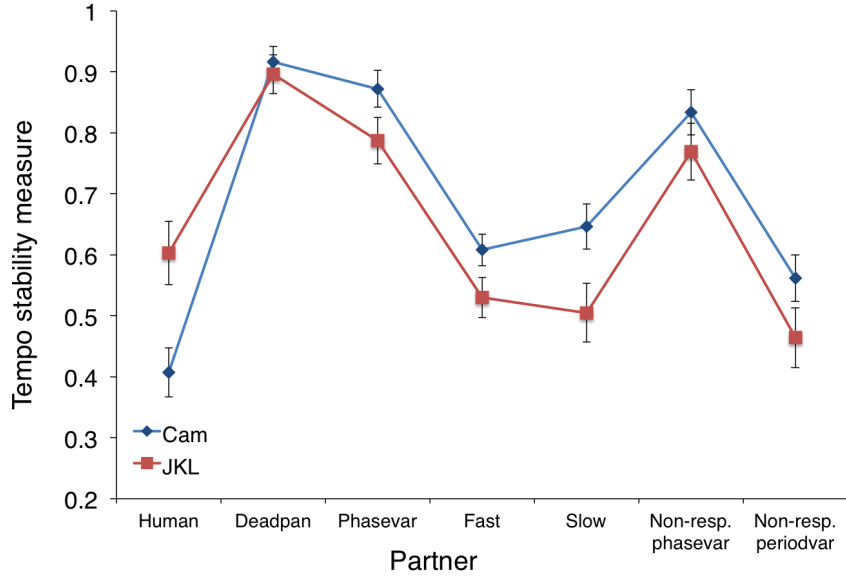


Figure 7.6: Tempo stability: Effect of *Partner* in the two *Groups*. Error bars represent standard error of the mean. Note the range of y-axis from 0.2 to 1.

the participant following the computer partner keeping accurate tempo. It is interesting to see that in the *Human* trials, where real interaction and mutual adaptation between the two participants was possible, tempo stability very low, even though the pairs synchronised best in this condition.

All the ANOVA effects are compiled into table 7.3

7.5 Discussion

This study extended the traditional finger tapping paradigm (Repp, 2005b; Repp & Su, 2013) to a two-person setting. The main aim was to compare entraining with another person to synchronising with computer tappers. Initially, it was hypothesised that tapping with a perfectly predictable, metronomic *Deadpan* partner would yield the highest tempo stability, as well as highest scores on the synchronisation measure. Partly for this reason, the *Phasevar* computer partner, which mimics human variability, was created, to be a closer comparison with the actual human interaction. This prediction

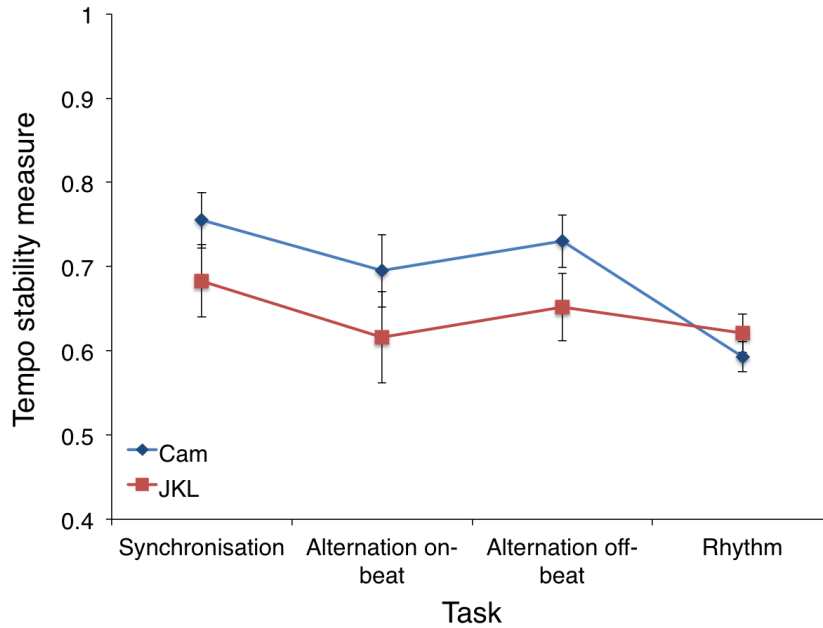


Figure 7.7: Tempo stability: Effects of *Task* in the two *Groups*. Error bars represent standard error of the mean. Note the range of y-axis from 0.4 to 1.

held in terms of tempo stability, in that synchronising with the computer tappers with a constant tempo helped to maintain the original tempo very well, while tapping with another human resulted in a tempo drift that was as large as when tapping with the computer tappers that were purposefully speeding up and slowing down.

In terms of synchronisation, the *Human* condition yielded the highest scores. It seems that the mutually adaptive entrainment between participants that is available in this context fosters both stability and keeping together with the other participant. Mutual adaptation of timing in interpersonal interaction has been noticed in a number of recent dyadic studies. Konvalinka et al. (2010) compared solo tapping to uni-directionally coupled tapping (one participant could hear the other but not vice versa) and to mutually coupled dyadic tapping (both participants could hear each other). Their results show that in the uni-directional case, the person who can adapt to their partner does so at a lag of one beat. Surprisingly, in the

Table 7.3: Summary of effects in ANOVA

<i>Source</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Synchronisation				
Task	2.2, 39.5	3.849	.026 *	.176
Partner	5.4, 51.6	23.637	.000 ***	.568
Group	1, 18	13.958	.002 **	.437
Task * Partner	7.3, 131.7	2.863	.007 **	.137
Task * Group	2.2	2.130	.128 N.S.	.106
Partner * Group	2.9	10.034	.000 ***	.358
Task * Partner * Group	7.3	2.606	.014 *	.126
Tempo stability				
Task	2.4, 47.9	5.300	.006 **	.209
Partner	3.8, 76.7	65.153	.000 ***	.765
Group	1, 20	1.029	.322 N.S.	.049
Task * Partner	7.0, 140.4	10.216	.007 **	.338
Task * Group	2.4	2.043	.132 N.S.	.093
Partner * Group	3.837	8.359	.000 ***	.295
Task * Partner * Group	7.022	2.727	.011 *	.120

Degrees of freedom adjusted with Greenhouse-Geisser correction.

Significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

case of mutual coupling, both participants adapt to each other's timing at a one-beat lag. This can be seen in cross-correlations of the inter-tap intervals, where both the -1 and +1 lags yield positive correlations. The zero lag correlation, however, is negative. In this study, instead of lags, I have analysed the amount of synchrony, which has allowed me to compare different uni-directional couplings to the mutual coupling condition. My results are consistent with Konvalinka's findings, and add that with mutual coupling, it is possible to reach very high entrainment even in the absence of steady tempo.

These results are a strong indication that the "information-theoretic" concepts and experiments on time-keeping and synchronisation only cover a limited range of the spectrum of human timing and entrainment abilities. In their investigation of tapping with adaptive partners, Vorberg (2005) and Repp and Keller (2008) studied how error correction gain (alpha coefficient in the linear error correction equation) influences synchronisation. In their analysis, they have concluded that in an ideal error correction, the sum of the gains (or the proportion of the perceived error that is corrected by the

next beat) should be close to 1. This corresponds with the description of mutual adaptation in this and in Konvalinka’s study. The linear model gets very cumbersome very quickly as the number of participants increases, and some of its assumptions (e.g. error correction parameters remain constant even across trials; timing is influenced only by the previous beat timing, etc.) do not sound convincing, but this aspect corresponds with these findings at least at a superficial level. However, as discussed in chapter 4.2, working out these linear model parameters from actual two-person data is complicated (Himberg, 2011).

This tapping study demonstrated that mutual adaptation (entrainment) can lead to very high degree of synchrony even simultaneously with high variability in tempo. This was unexpected, but this result is in line with findings in other domains that illustrate the “power” of mutuality. For example, in a mirror game, participants reach “co-confident motion” during joint leadership conditions and better entrainment than when there is a designated leader (Noy et al., 2011). Also, in synchronising speech, the best results are reached when speaking in a live interaction than when trying to synchronise with a recorded version (Cummins, 2009). Nevertheless, further studies of especially the flexibility of timing and synchronisation in dyadic and group settings are necessary. For example, in this study the two groups had different baseline tempi (120 BPM vs. 100 BPM), and the slower group turned out to achieve lower synchronisation scores in all conditions except when tapping with another human. Mutual adaptation seemed to cancel the effect of increased noise in a slower tempo. This selective effect of tempo should be investigated in a specialised study. Also, the role of auditory and visual channels or matching metrical schemata could be investigated. I have carried out a few pilot studies on some of the variables; these are discussed briefly in appendix A. The experiments in the next chapter also are in part based on these findings, in that they acknowledge the need to study timing and entrainment in natural settings, and focus on the relative timings between participants rather than absolute timings or a perfectly isochronous referent. They expand the traditional synchronisation task by including actual music-making and take a step towards embodied cognition research by analysing continuous body movements rather than discrete tap onsets.

Chapter 8

African dance

8.1 From social to social embodied cognition

As was shown in the previous chapter, adding the social dimension into an experiment was technically a relatively small change, but it caused unexpected results. The surprising main finding was that participants reached a very tight synchrony in human-human pairs even when their tempo was drifting a lot. Essentially, pairs of participants were “drifting” as much as when tapping with a computer tapper that was deliberately speeding up or slowing down, and yet they managed to synchronise with each other better than with a perfectly predictable, metronomically precise computer tapper.

The next step is to include an analysis of full-body movements, in order to investigate entrainment and embodied and social cognition, we conducted two studies on African dance. We used optical motion capture (see below and appendix B for more details) to record three-dimensional movement data from the participants. In both studies, we were especially interested in the rhythmic entrainment of dancers, the embodiment of metre of different metrical layers, as well as the role of familiarity (as a proxy for cultural background) on these factors.

In these studies we have European and African participants in essentially cross-cultural study settings. As mentioned above, drawing exact connections between their cultural and musical backgrounds to the performances in these tasks is not unproblematic, and would require a more detailed analysis. Therefore we decided to use the terms “expert” and “novice”, as they capture what we think is the main factor differentiating these groups.

8.2 Choir study

8.2.1 Choir study: background

The first study focused on choir singing, which is one of the world's most common musical activities (Nettl, 2000). Choir singing is shown to promote social cohesion and have other health benefits (Beck, Cesario, Yousefi, & Enamoto, 2000; Kreutz, Bongard, Rohrmann, Hodapp, & Grebe, 2004), and is a cognitively complex activity that recruits cognitive resources from memory to voice control, from auditory perception to aesthetic and social processes (Gabrielsson, 2003; Palmer & Meyer, 2000).

Choir performance also poses high requirements for interpersonal synchronisation and coordination of movements and vocal outputs. As in any ensemble performance, very high levels of entrainment and coordination must be achieved, and strict aesthetic criteria in terms of the quality of vocal outputs must be met (Grafton & Cross, 2008; Janata & Grafton, 2003; Rasch, 1988; Sawyer, 2006). Pitch, timing and timbre need to be controlled at individual and group levels, meaning that participants must divide their attention between their own part, the parts of the others, and the emerging whole (Keller, 1999). Rhythmically, the individual performance needs to be adapted to those of the others in a co-operative, mutual fashion.

Studying rhythmic entrainment in choir performance solves many of the issues of ecological validity and musical fidelity that plague finger tapping studies. At the same time, however, the complexity of the behaviour means that experimental control is much more difficult than in tapping studies, and similar focus on individual variables might not be possible. However, in accordance with the interactionist approach discussed in chapter 2.6, the focus is shifted to the interaction of the participants instead.

We had the opportunity to study performances of Emmanuel Lutheran Choir (ELC) from South Africa, joined by a group of Finnish choir singers. We organised a two-day workshop where the South African experts would teach the Finnish novices their songs and dances, and in return, the Finnish participants would teach one traditional Finnish song to the South Africans. During the course of the workshop, a number of motion capture sessions were conducted. We also recorded audio and video, and primarily wanted to study the embodiment of rhythm and metre and analyse the mutual synchronisation of the performers' body movements. We were also interested in

how familiarity with the material influences entrainment and how it unfolds in time.

Having experts and novices performing together, we were hoping to see the effects of familiarity on synchronisation accuracy, and also whether familiarity would be reflected in how entrainment unfolds in time, or the dynamic aspects of it: as in the cooperative tapping study, we were interested in seeing whether the synchronicity is a result of mutual adaptation or whether the experts act as leaders that the novices follow.

8.2.2 Choir Study: participants and materials

This study was conducted in conjunction with a cross-cultural choir workshop in Jyväskylä. The Emmanuel Lutheran Church Choir from Johannesburg was visiting Finland on their tour, and we recruited Finnish choir singers from Jyväskylä to join them in the workshop. The repertoire of the Emmanuel Choir consists of traditional African music, originating mainly from various southern African cultures (e.g., Zulu, Sotho and Xhosa). As their performances combine singing with coordinated movement choreographies, they were optimal for studying movement coordination in groups, a topic that had interested us already for some time.

While the workshop provided us with a good opportunity to record interesting data, the ad hoc nature of it meant that we could not construct a proper, balanced cross-cultural experiment (Berry et al., 1997). Rather, we focus on a more descriptive level of analysis, and as discussed in the introduction, the crucial difference between the two groups of participants is their expertise and familiarity with regards to the musical style and material rather than their background as such.

In terms of musical expertise in general, the groups were similar, as the South African participants were actively performing professional and semi-professional musicians, while the Finnish participants also had musical training, all performed in their choirs regularly, and the group included professional musicians. The South Africans were, of course, experts in the musical styles that they performed in, and also familiar with Western popular and classical music, but not familiar with the Finnish or European folk genre that the Finnish song in this experiment represented. Similarly, while the Finnish participants had some limited exposure to African music cultures, they were not familiar with the particular genres they learned in this

workshop.

During the two-day workshop, half a dozen African songs and one Finnish song were learned and performed. One South African and the Finnish song were chosen for this analysis. The South African song was *Fiela Ngwanana*, (*“Sweeping Girl”*), a Sotho song. The Finnish song was *Kukkuva kello* (*“Cuckoo Clock”*), a folk song familiar to all Finns that is often sung as a canon. As the Finnish song does not usually come with a choreography, the South African participants prepared one for it. So the movement style of the Finnish song was similar to the African songs, but simpler.

8.2.3 Choir study: apparatus and procedure

During the workshop, a selection of song performances were recorded using optical motion capture (infrared cameras, see Appendix B). Normal video and audio recordings were also made. For each of the captures, a smaller subset of the 40+ workshop participants were selected based on volunteering. These volunteers wore reflecting markers and performed in the centre of the room, while the rest moved away from this “capture volume” to the sides of the room. In the performances, all workshop participants still contributed to the performance, although the movements of only a small group were being captured.

In the African song, 8 participants (4 expert, 4 novice) wore markers in their heads and feet. They were positioned at the centre of the room in a 2 by 4 –formation. The participants (10, 7 expert, 3 novice) of the Finnish song were positioned into a 2 by 5 -formation. In the Finnish song capture, the participants only wore markers in their heads, as the larger number of people in the same small space would make the feet markers not visible to the cameras.

The songs were learned aurally, i.e. by rote, and the choreographies by imitation. For each of the songs, the process was similar, and the individual parts were learned one by one. This is the typical way of teaching folk songs in both countries, although the Finnish group was perhaps less used to it and more used to learning music via notation.

8.2.4 Choir study: data analysis

Movement data was processed and analysed using MATLAB and the MoCap Toolbox (Toiviainen & Burger, 2010). We focused on pairwise and group entrainment, first comparing the experts and novices using pairwise cross-correlation, looking at the vertical acceleration of the feet in the African song and the vertical and horizontal acceleration of the head marker in the Finnish song. We calculated the cross-correlation for all the pairwise combinations within and between the two groups.

The temporal development of synchrony is essential, as discussed in the previous chapters. To this end, we conducted windowed cross-correlation analyses on the vertical movements of the head markers, as these markers were used in both songs and they capture the overall periodicity of body movements very well (Toiviainen et al., 2010).

To investigate entrainment at the group level, we can either look at pairwise entrainment measures as above, but eventually proper group entrainment analysis methods are needed, as well. We used the group synchronisation model developed by K. Kuramoto (Acebrón et al., 2005). Looking at phase coherence in the group of oscillators, an order parameter can be calculated, and this can then be used as an index of entrainment within the group (it ranges from 0 to 1 and is conceptually very close to the \bar{R} that we introduced as a pairwise index of entrainment in chapter 5. Details of this procedure can be found in section 6.8 on group entrainment).

For this data, before estimating the instantaneous phase of each of the head markers, we first filtered the position data using a Gaussian filter, using the tactus of the song as the center frequency. In these choreographies, steps and most salient movements tended to co-occur with the beats at the tactus level, making it a fruitful frequency band to look at in terms of group entrainment. This filtering of course removes information from the data, but it allows the instantaneous phase of the movement to be calculated unambiguously using the Hilbert transform (Khvedelidze, 2002). From the phases of the group's movements, the order parameter and individual deviations of the group phase could then be calculated.

8.2.5 Choir study: results

Pairwise analyses

Cross-correlations of the vertical acceleration of the foot markers in the African song and vertical and horizontal acceleration of the head markers in the Finnish song were calculated, up to a lag of 0.5 seconds, about the length of one inter-step interval. All possible pairwise combinations of participants in each group were analysed, so that the entrainment of each group member with each other group member were considered. The maximum coefficients and the lags at which they occur are presented in figures 8.1, 8.2 and 8.3. Figure 8.1 shows the group differences while individual differences are presented in figures 8.2 and 8.3.

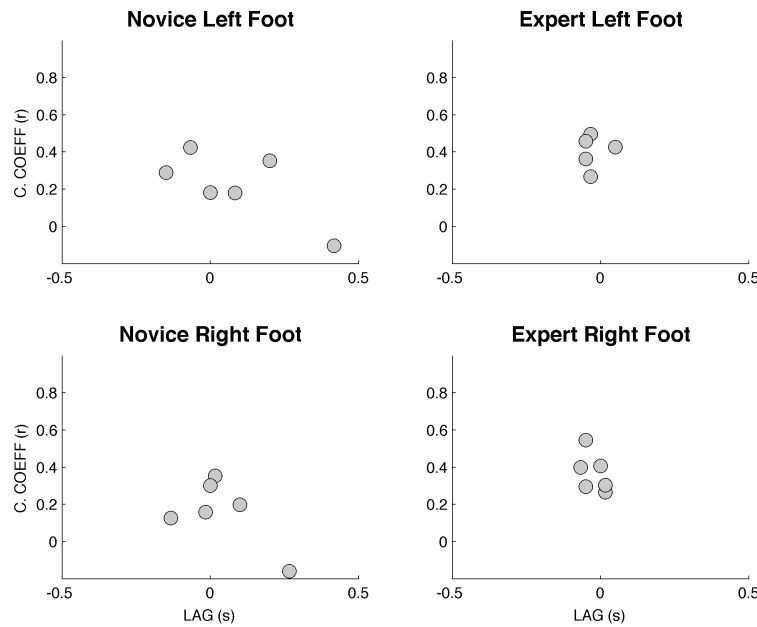


Figure 8.1: Pair-wise cross-correlations for foot acceleration in the vertical dimension in the African song, displayed by group. Circles represent the highest correlation coefficients (r) between the six possible pairings of participants within each group. Y-axis represents the strength of the correlation and x-axis shows the lag at which the maximum correlation occurs.

In figure 8.1, the circles in each of the subplots represent the maximum

correlation coefficient (r) between the six possible pairings of participants within each group in the African song. Instead of comparing each of the pairs to each other, we focus on the group level synchronicity here. The correlations are relatively weak, which suggests that participants executed the steps in this part of the choreography in a number of different ways; they were not supposed to march in lockstep but rather lift their feet in time with the music. Another factor that lowers the correlation is that we used unfiltered data here (e.g. rather than band-pass filtering it to only contain frequencies close to the step period). However, the expert group seems more coherent than the novice group, not only are their maximum coefficients slightly larger, but also they tend to occur at lags closed to zero.

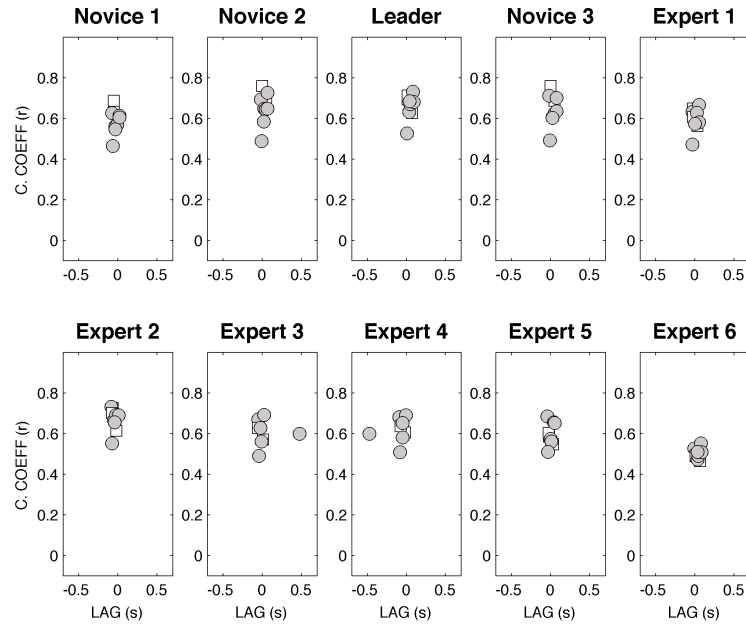


Figure 8.2: Pair-wise cross-correlations for head acceleration in the vertical dimension for the Finnish song. Each subplot shows the maximum correlation coefficient between an individual and each of the other performers. Shaded circles indicate correlations with experts and white squares indicate correlations with novices. The x-axis shows the lag or synchronization error between pairs and the y-axis shows the strength of the correlation between pairs. The subplot entitled ‘Leader’ is for the expert participant acting as leader of the performance.

Figures 8.2 and 8.3 show the data for cross-correlations between performers in the Finnish song, in vertical and left-to-right dimensions, respectively. As in figure 8.1, the maximum correlation coefficients and their lags are plotted, but this time for each individual separately. For example, the top left subplot shows coefficients between Novice 1 and all other performers. The correlation coefficients are much higher in the vertical dimension. The choreography in this section was based on walking along the horizontal dimension (walking left to right and back). To be synchronised in the vertical dimension, all you needed to do was to get the timing of steps right, but to get the horizontal dimension synchronised, you would also need to get the amount and timing of this lateral movement in sync. According to these results, the step timing was much better achieved than timing of the lateral movement.

Comparing these plots to video recordings from the performance reveals important properties of the group performance. The video shows that Novice 1, at the front left corner of the formation, is behind the rest of the group in her movements in relation to the choreography. The movement data reveals that in terms of the timing of the individual steps, as such, she is in time with the others. This can be seen by looking at the vertical dimension. However, in terms of her lateral movement, she stands out as being late in relation to the others.

8.2.6 Choir study: examples of windowed analysis

Each data point in the analysis above summarised a whole trial section. To reveal the temporal development of the synchronicity in these pairs, a windowed analysis is necessary. Instead of investigating all pairwise combinations of participants, we chose the front-left pairs (novice and expert) for this analysis.

In these cross-correlograms (figures 8.4 & 8.5), darker shading represents higher correlations, the vertical axis represents lag, and the horizontal axis time. If the darker shades occur below the zero-lag line, the expert is leading, and dark shades above the line suggest that at those points in time, the novice is ahead.

The highest correlations seem to occur very close to the zero-lag line, which indicates that the two participants are synchronised without a major lag. Also, as the peak correlation fluctuates between being at a positive and negative lag, it suggests that neither of the participants has a clear leadership

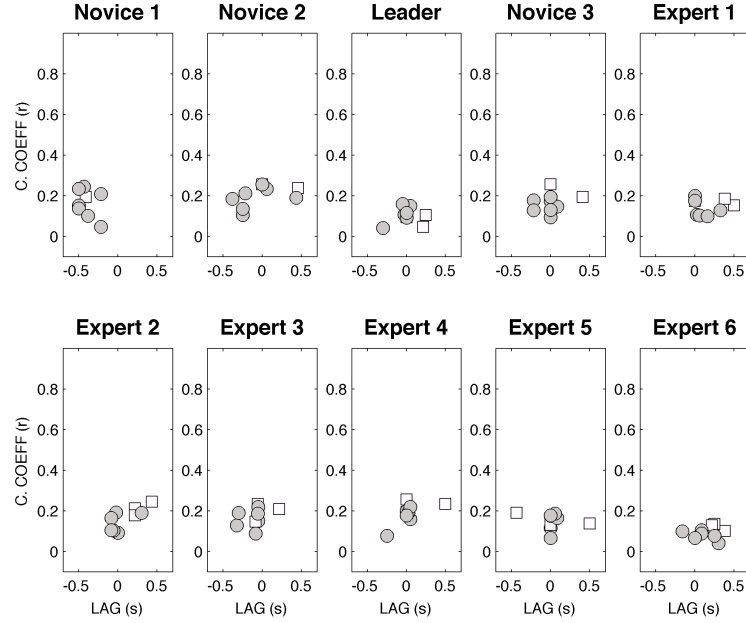


Figure 8.3: Pair-wise cross-correlations for head acceleration in the horizontal dimension in the Finnish song. Each subplot shows the maximum correlation coefficient between an individual and each of the other performers. Shaded circles indicate correlations with experts and white squares indicate correlations with novices. The x-axis shows the lag or synchronization error between pairs and the y-axis shows the strength of the correlation between pairs. The subplot entitled Leader is for the expert participant acting as leader of the performance.

role, rather the two continuously and mutually adapt to each other in order to achieve the synchronicity of their movements. The movements are periodic, and thus the sidebands of high correlation appear at plus-minus the period of the movements. In the Finnish song the rhythm is un-syncopated, with the rhythmic and metrical accents aligning perfectly, amplifying this effect.

Choir study: phase synchronisation in the group

In addition to looking at how the two groups differ in terms of their coordination, it is interesting to see how closely the whole group entrains. As described in the methods section (8.2.4), the order parameter is calculated

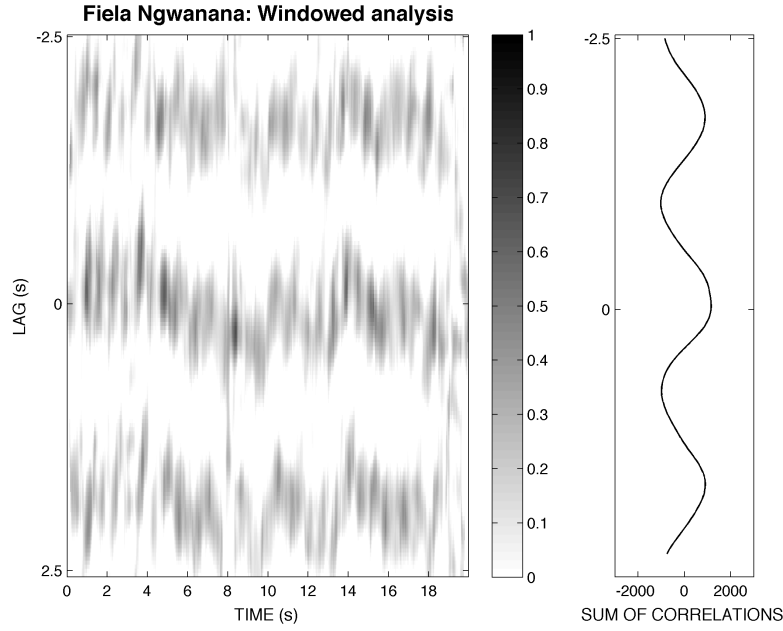


Figure 8.4: Cross-correlogram and sums of correlation values for each lag, vertical position of head for one pair of expert and novice performers in the African song. Darker shades indicate higher correlations.

by first extracting the instantaneous phase for each participant and then comparing them to each other. The order parameter is a measure of phase coherence in the group, and serves as an index of entrainment. Plotting the order parameter over time reveals how group coherence evolves in the performance.

Figure 8.6 shows that the order parameter gets very low values in the beginning and the end of the performance, but rises to a respectable level for the midsection of the song. This reflects the choreography of the song, where the beginning and end of the song are relatively motionless, and the step pattern starts only at the beginning of the second verse. For consistency, in this figure, the Finnish participants are labelled as novices even though they are more familiar with the song and style of music than the African group. The Finns are the more coherent group here apart from occasional dips in the order parameter. As there are only three Finns and seven South Africans in this performance, we also compared the in-group order parameter

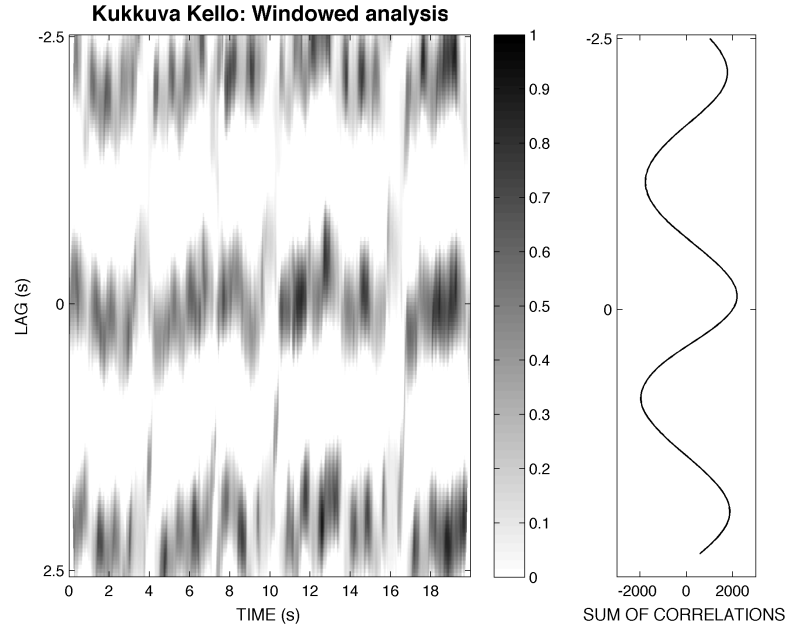


Figure 8.5: Cross-correlogram and sums of correlation values for each lag, vertical position of head for one pair of expert and novice performers in the Finnish song. Darker shades indicate higher correlations.

of the Finns to a randomly selected groups of three South Africans, but the Finns are still the better entrained group. Thus, it is not merely a function of a smaller group being more easy to coordinate (the Finns were dispersed around the formation and not standing next to each other), but it seems to reflect a genuine difference in expertise and/or familiarity with the song.

In this analysis, the data was band-pass filtered at the frequency of the tactus, and thus the order parameter reflects the group entrainment at that metrical level. However, this is of course not the only level at which the participants' body movements can be entrained. The lateral movement, for instance, has a longer period of approximately 3.5 seconds. We conducted the same analysis on the horizontal data, filtered at that frequency and obtained similar results.

Plotting the order parameter for the South African song yields a different curve but again the nature of the choreography is reflected in the results (figure 8.7). The order parameter marks the boundaries between different

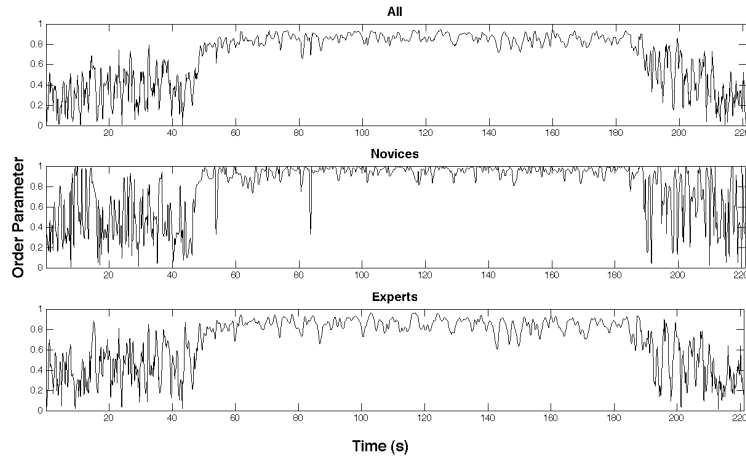


Figure 8.6: Order parameter for the whole choir, the novices and the experts, for the Finnish song.

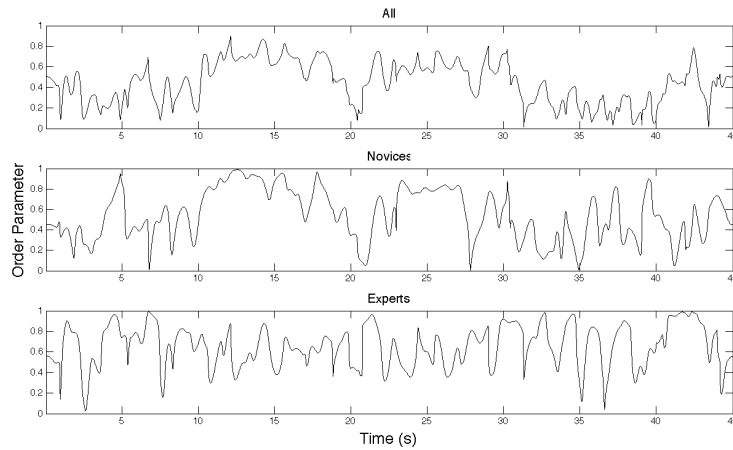


Figure 8.7: Order parameter for the whole choir, the novices and the experts, for the African song.

segments of the choreography, with lower values at the segment boundaries, especially for the novice group (the Finns). Probably the experts had fewer problems remembering what the movement pattern in the next segment is, and thus could perform the transitions in a more coordinated fashion. It is notable that the songs were practiced segment by segment, and thus the

novices had focused on remembering how each segment was performed, but transitions from one segment to the next were not explicitly rehearsed.

Figure 8.8 depicts the individual deviations from the average or group phase in the performances. This could be used to evaluate individual roles in the group. For example, in the South African song, the whole group is relatively closely entrained, apart from one novice who lags in relation to everyone else. In the Finnish song, the individual deviations are larger in general. Participant 3, given the role of the leader in this performance, is leading also in the average phase deviation.

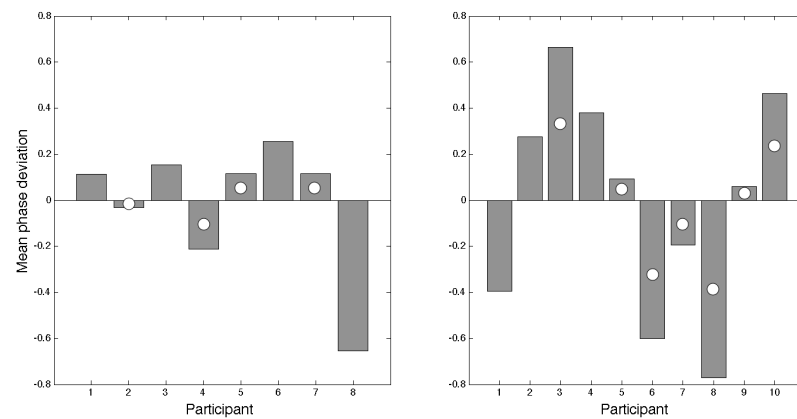


Figure 8.8: Deviations from the mean phase for each of the participants. African song in the left panel, Finnish song in the right panel. White dots indicate experts.

Choir study: metrical levels

We did not conduct a systematic investigation of metrical levels and body movements like Toiviainen, Luck and Thompson (2010) did, as in our study, the movements were choreographed and not spontaneous. We could thus rely on studying entrainment at certain individual metrical levels, as in the previous section. However, we stumbled upon evidence that corroborates their findings that different metrical levels are embodied in the movements of different parts of the body. In figure 8.9, the vertical foot and head movements of two participants are depicted. These plots represent five second sections

of the South African song. For the novice, the foot and head trajectories have the same overall rate, and the head movements closely follow the foot movements, although the head movement contour is simpler, as feet make an additional dip mid-beat. Essentially, when she takes steps, the whole body moves at that rate. In contrast, the expert's head movements are more complexly coupled to the foot movements. While the foot movements are identical to the foot movements of the novice, showing that they take the same steps, the head trajectories are very different. The head trajectory of the expert seems to be a summation of up and down movements at several different metrical levels, there is the faster beat level movement as in the foot movements, but also the slower oscillation at the bar level.

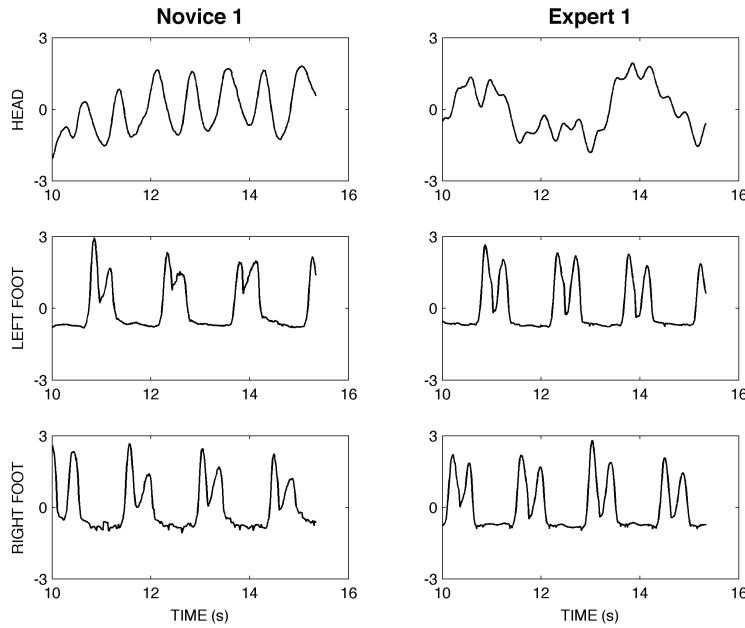


Figure 8.9: Vertical head and foot acceleration, comparison of the expert and novice performers (the pair in the right of front row) in one section of the African song. For the purposes of presentation, the data was z-scored and the y-axis shows standardized values.

The vertical movement of the head marker thus seems able to capture multiple metrical levels at once, which is good news for motion capture using a very small number of markers. Also, it suggests that perhaps the experts

had a richer metrical schemata or at least they were able to embody more levels than the novices, who were highly focused on the tactus level and taking the correct steps, while remaining quite stiff in doing so.

8.2.7 Choir study: discussion

The opportunity to run the Choir Study presented itself rather surprisingly, which explains the ad hoc nature of the study. We did not know the choir, or its members beforehand, and thus we were not sure what kinds of experimental manipulations we would be able to do; for example, whether we could ask them to perform just the movements without the singing and vice versa, or if it would be possible to break the choir down to smaller groups for recordings. Also, we were not sure how easy or difficult the songs would be for the Finnish participants to learn, and whether we could measure groups of Finns and groups of South Africans separately or in mixed groups. In the end, the Emmanuel Choir members were extremely helpful and flexible, and the Finnish participants quick to learn and equally gracious and eager to oblige to our wishes and ideas. We decided to take a descriptive approach and collect data that could be used in group entrainment analysis as well as tracing the novices' learning process.

Previous studies on entrainment had mostly looked at dyads, and the few that have looked at small ensembles have focused on pairwise entrainment. This was our approach as well at first, as we measured the pairwise cross-correlations between all novices and all experts in the Choir Study. Plotting the maximum coefficients and their lags gave us a compact and intuitive visualisation of the coherence of these groups. The experts were the more coherent of the groups perhaps because their metrical schemata were more similar to each other, while the novices all tried to grasp the metrical and rhythmical patterns of the movements in their own fashion.

In order to investigate the temporal dynamics of the pairwise interaction in the group, we conducted cross-correlation analyses in a moving window. For a novice-expert pair, this analysis revealed that the overall entrainment in the pair was a result of continuous, mutual adaptation. The pair was entrained at a zero lag, which in itself is unsurprising, as they both were dancing and singing the same tune. However, given their very different levels of expertise, one could also expect them to have different roles, for example those of a leader and a follower. The analysis shows, however, that

they are much more equal in their mutual influence, both give and both take, each is ahead and behind a little at different times, as they are “negotiating” what the optimal tempo and microtiming of the movements should be. This mutuality has been seen in most dyadic studies (Goebel & Palmer, 2009; Konvalinka et al., 2010; Noy et al., 2011; Pecenka & Keller, 2011; Yun et al., 2012), including the one reported in chapter 7 in this thesis¹.

The windowed version of the cross-correlation has the limitation that it can only take two variables at a time. Therefore we used the Kuramoto model of phase entrainment to measure the level of order in a group. Musical movement with a clear periodicity provides good data for such analyses. First the data was filtered in order to focus the analysis on movements at the most salient metrical level. As our data was stationary, we could use Hilbert transform to obtain the instantaneous phase, and after that the estimation of the group order parameter was straight-forward. The order parameter serves as an index of entrainment in a very similar way as \bar{R} , the concentration measure for the circular distribution of phase angles. In this study, the order parameter was shown to be sensitive to different segments of the choreography, and it also indicated differences between the novices and experts in their group coherence. It was the Finnish group showing higher order when performing a Finnish song with African moves. Again, this was very likely a familiarity effect modulating the amount of attention that is left after trying to remember the words, melody and moves of an unfamiliar song (cf. Keller, 1999; Sawyer, 2006).

As the Finnish group had very little or no experience in performing (South) African songs or dances, the choreographies (chosen by the expert group from their concert repertoire) were relatively simple, yet the skill difference between the groups remained relatively high. In order to study the finer details of body entrainment, and possible contrasts between groups, we conducted an experiment where all participants were well-versed in African dancing.

¹Alternatively, this could be a conscious strategy on the part of the expert, to “guide” the novice by pulling and pushing in time. However, the shifts around the zero lag are very small and the fluctuations rapid, which would be difficult to control consciously.

8.3 Dance study

8.3.1 Dance study: background

In the other study, we focused on dyadic entrainment in Kenyan Ohangla dance. The study reported here is a pilot conducted with four participants, two Kenyan and two Finnish. The Ohangla dance is much more complicated than the choreographies in the first study, and thus even the Finnish “novices” in this study were experienced amateur dancers. In this second study, the methodology was very similar to the first one, but rather than recording over a longer period, all recordings were conducted in one short session. Also, a more complete full-body marker set was used, so that the body movements could be analysed in more detail. To look at effects of familiarity and cultural background, we formed Kenyan and Finnish as well as mixed pairs from our four participants.

Primarily, we expected to see an effect of expertise on the coordination and synchronicity of movements. Kenyan dancers should perform better, as their expertise on this style would allow for more attentional resources to spend on ensuring high levels of coordination (Keller, 1999). In addition, pairs with matching cultural background should perform with higher synchronicity than unmatched. These two factors are linearly combined into our hypothesis, as summarised in figure 8.10. Contrast analysis (Furr & Rosenthal, 2003) requires a very specific hypothesis against which it tests the data. We formed matching and mixed pairs from our four participants, and assigned weights for expertise and matchedness for each pairs, attempting to predict their performance. In effect, we hypothesise that the higher the level of expertise and the higher the matchedness, the higher the synchronicity or coordination in the pair.

The study presented here is a pilot, as with such a small sample of participants, no firm conclusions about the issue can be made. The main aim of this pilot study was to develop a research paradigm for studying embodied metre in dyadic settings, especially in the naturalistic context of skilled dance moves. Therefore we did not conduct a thorough musicological analysis of the music and dance involved, nor did we dig very deeply into the musical background of our participants. Thus we are unable to conclude whether any differences that are found can be attributed to factors in the metrical organisation of the music, and/or the participants’ musical backgrounds. Also, in

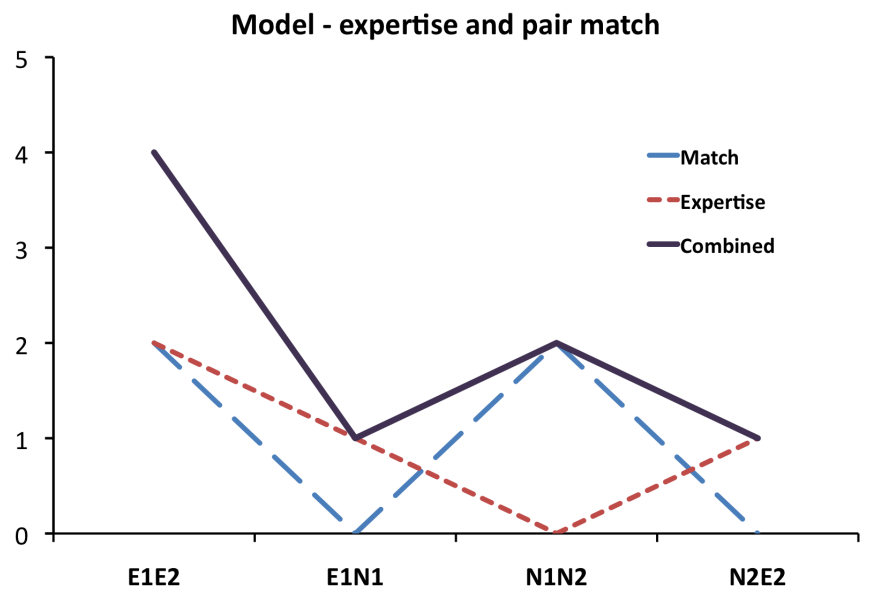


Figure 8.10: Hypotheses - how expertise (dashed blue line) and matching expertise (dotted red line) are combined to predict synchronisation accuracy (solid purple line).

a larger study, musics from both target cultures should be used, and possibly control conditions, in order to get a clearer picture of these effects.

8.3.2 Dance study: methods

Dance study: participants, materials, and procedure

We recruited four participants (all female) for this study, two Kenyan experts in Kenyan Ohangla dance, and two Finnish novices. All were students at the Music department of the University of Jyväskylä at the time. The two Kenyan students had studied the Ohangla dance as part of their music studies at Kenyatta university. Both Finns had dance training and had limited prior experience in African dancing, but they were not familiar with this style. The choreography was selected by the experts, and taught to the novices in a 35-minute session prior to the start of the experiment.

The choreography was 90 seconds long and was performed to a drum accompaniment, which in our experiment was recorded earlier in a workshop

and provided through loudspeakers. The choreography consists of a series of segments, and within each segment, a movement pattern repeats a number of times. Both these features make this choreography well-suited for synchronisation analysis. The music is strongly rhythmic, and as is typical for Kenyan drumming, it has polyrhythmic patterns that are embodied in the dance through different parts of the body (Senoga-Zake, 2000). Also, the repeating patterns mean that also longer periodicities could be investigated.

A total of eight performances were recorded. Each of the four participants first performed the choreography solo, after which four different pair performances were recorded: Expert1-Expert2 (E1E2), Novice1-Novice2 (N1N2), Expert1-Novice1 (E1N1) and Expert2-Novice2 (E2N2).

Dance study: data analysis

Each participant wore a full-body marker set of 28 reflective markers, placed on the joints and major body parts according to figure 8.11. We focused our analysis on the body parts marked with red dots, the head centroid, hip centroid (root) and the right hand. These were selected because they had prominent roles in the choreography, and are important kinematically. The head and root represent the movement of the core of the body, and as mentioned above, the head marker often captures the different metrical levels of periodic dance movements. Right hand marker was important in this choreography, and it had the largest variance of movement out of all the markers. The head and root centroids were calculated as the mean position of the three head markers and the three hip markers respectively. The right hand was represented by the marker that was placed on the right hand middle fingers of the participants.

MATLAB and MoCap Toolbox were used for this analysis, as well. The analysis presented here concentrates on vertical movements in the pair performances, as we were mainly interested in entrainment. We selected one of the segments for analysis, and used the beats of the music to identify the start and end times of the segment. At first, we tried to use the movement itself to identify the boundaries of segments, but it turned out that minor asynchronies in the movements made it impossible; if a segment starts with a step, should it be the step of performer 1 or 2 whose step we use as a marker? Thus, the beat of the music was a better, more objective signal to use for marking off the boundaries.

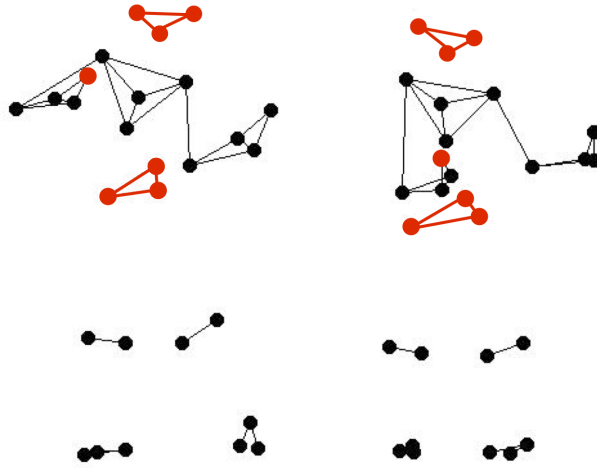


Figure 8.11: A motion capture image of the Expert pair performing the “double hip shake” -segment of the choreography. The optical motion capture system tracks the trajectories of the markers attached to the performers. The areas of interest are highlighted in red.

We termed this segment the “double hip shake”, as in this movement the dancers take sidesteps left and right in alternation, and make an extra hip wiggle at the last step to each direction. In this segment, the participants are standing side-by-side, which facilitates data analysis, as the two participants’ movements have the same directions along the horizontal dimensions without having to do coordinate conversions. Also, the repetitive pattern in this segment is relatively simple and it repeats many times, and the segment itself repeats in the choreography twice. Thus, the first and second occurrences of the segment can be used as repeated measures, and the movements in the segment itself should afford the highest possible levels of synchronicity, as the participants perform many repetitions of the pattern.

8.3.3 Dance study: results

Dance study: windowed cross-correlation

Looking at the temporal evolution of the pairwise entrainment in the Dance Study shows a similar novice-expert difference as the movement patterns in

the Choir Study. As the windowed cross-correlations in figure 8.12 show, the novices seem to entrain at a frequency double that of the experts. The mixed pairs yield mixed results, but overall the entrainment remains very stable in all pairs.

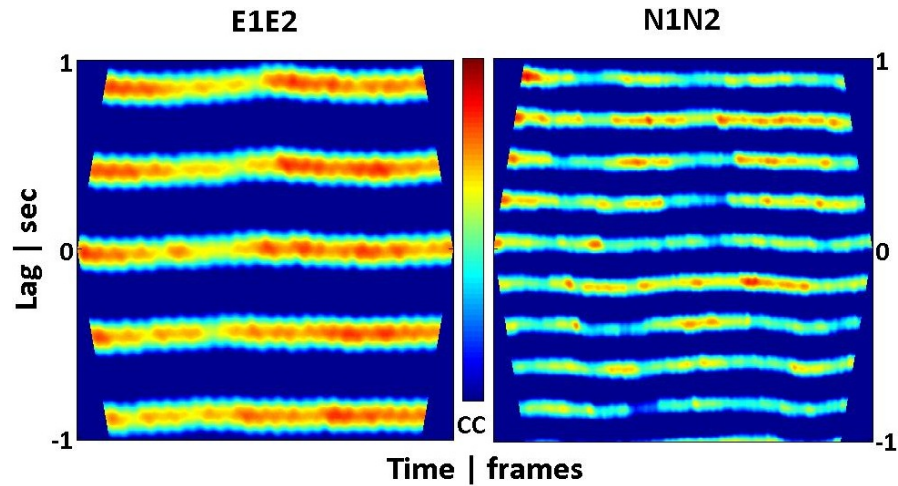


Figure 8.12: Windowed cross-correlations of head centroid velocities in vertical dimension.

Dance study: bounding rectangle

There are many ways to measure the overall amount of movement from the movement data, for example, a commonly used one is to look at the cumulative distance travelled by a marker or a set of markers. However, this measure does not take into account how large or small the movements are, and jitter of the markers (noise due to small measurement errors in the system) might accumulate a lot of distance for a marker even when it is almost stationary (Jensenius, Nymoen, Skogstad, & Voldsund, 2012). A method that takes into account the range of the movements and not just the amount of movement is to measure the bounding rectangle for the movement. For example, the bounding horizontal rectangle is the smallest rectangular shape that contains the markers in the two horizontal dimensions during the given time window. For the dance data, we calculated this horizontal, average bounding rectangle (ABR) squared, so that it looks at the speed of displacement. The calculation was done in a four second window, with

consecutive windows overlapping by two seconds. The results indicate that in general, the novices had a larger speed of displacement than the experts by 27% (Novices $ABR = .007 \text{ m}^2/\text{s}$; Experts $ABR = .009 \text{ m}^2/\text{s}$). This was based on the root, head and right hand markers for all dance performances.

Dance study: phase synchronisation

We used Hilbert transform on the vertical position data to calculate the instantaneous phase for each participant, segment and marker. The phase difference between the two participants was then calculated (modulo 2π), and the concentration and mean direction of this distribution of phase differences were then calculated. These are the entrainment measures \bar{R} and θ that we are already familiar with. Figure 8.13 shows the rose plots or circular histograms of the phase differences for the head markers.

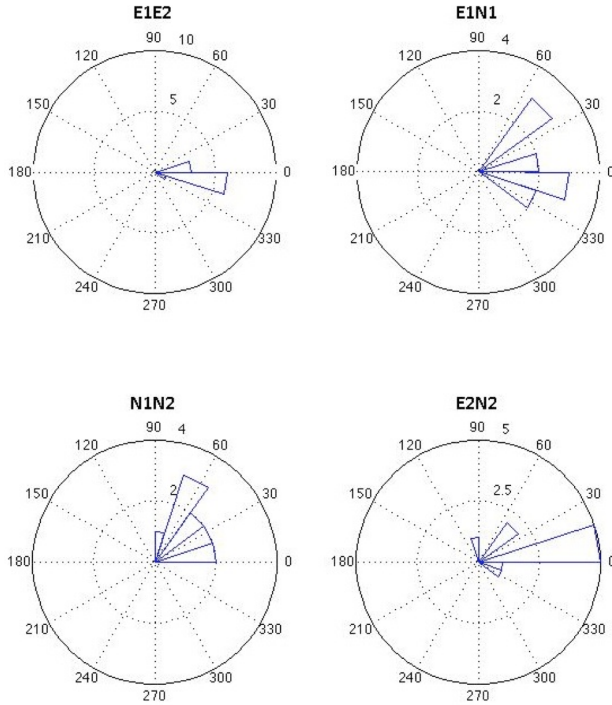


Figure 8.13: Rose plots, or circular distributions of the phase errors of the vertical position of the head markers.

In comparing these four performances note that the four subplots have

different ranges of the value axes. The expert pair is the best entrained, while also the second mixed pair has a mean direction close to zero. The novice pair seems to be entrained at a 90 degree phase difference, and in general their phase error distribution is wider and has a less pronounced peak. It is worth noting that these are head movements, and footsteps or hand waves would be better synchronised than this, or the phase differences would be easier to notice. The head movements in this choreography are accentuating the relatively strong vertical torso movements, and each dancer could have different ways to implement this movement (e.g. either as a movement that is in phase with the torso movement, or as being slightly delayed in relation to it, as a sort of a whiplash motion, a wave that travels from the torso to the head).

Dance study: contrast analysis

As was discussed in the methods section, the high dimensionality of movement data is often reduced in the analysis stage to gain in clarity. Often, the movement of markers is condensed into individual numbers that capture a feature of the movement. This could lead to problems regarding statistical power. Using traditional “omnibus” methods such as analysis of variance (ANOVA), leads to a further loss of statistical power as unspecific hypotheses are being tested (Furr & Rosenthal, 2003). Contrast analysis is a method where hypotheses are clearly specified and even quantified before any testing occurs, so that the data needs to be checked against only one, specific hypothesis, and the statistical power can be maximised.

Our two hypotheses were plotted in figure 8.10. We hypothesised that two factors, matching cultural background and “raw” expertise would lead to higher synchrony. These two hypothetical factors are linearly combined to yield a set of weights ($w_h = [4, 1, 2, 1]$), which are then normalised: ($w_{hnorm} = [0.8165, -0.4082, 0, -0.4082]$). The measured \bar{R} results are then compared to these weights using a simple one-way t-test. According to this analysis, the data fits the hypothesis extremely well. In entrainment, the expert pair is the best, followed by the novice pair, followed by the mixed pairs. This fits the pattern of our prediction. For the root centroid, the t-test yields $t = 8.4817, p = 0.0374$, with the effect size estimate of $r = 0.9931$. For the head centroid, the effect size estimate is even higher, $r = 0.9944$. The hypothesis fits the data extremely well, and the two factors correlate

Table 8.1: \bar{R} -values of the pair performances - head and root centroids and right hand marker.

		E1E2	E1N1	N1N2	E2N2
Root	Segm 1	0.95	0.52	0.84	0.44
	Segm 2	0.84	0.70	0.78	0.23
Head	Segm 1	0.93	0.54	0.68	0.53
	Segm 2	0.76	0.53	0.54	0.34
Hand	Segm 1	0.91	0.35	0.37	0.44
	Segm 2	0.73	0.88	0.65	0.81

almost perfectly with the measured synchronicity. The results are collated into table 8.2. The set of 5 contains the three markers that have been used in the analysis so far, plus the upper torso centroid and the right ankle. Thus it represents the body as a whole.

Table 8.2: Effect sizes of the contrast analysis.

	Root	Head	Hand	Set of 5
Hypothesis	0.9931	0.9944	0.5270	0.7981

These results suggest that the further from the core of the body we look, the less clear the entrainment becomes. For example, the head and the root are very closely synchronised, and match the model very well, while the hand movements have more artistic freedom and allow for more individual expression. Thus our simple model does not capture the differences in hand movement synchronies between the different pairs.

8.3.4 Dance study: discussion

The four-person sample presented here forms a pilot run for a future experiment. In our pilot, windowed cross-correlation analysis showed that the experts entrained at a higher (slower) metrical level than the novices. Also, we observed that pairs matching in expertise reached higher levels of entrainment than mixed pairs. This is an interesting result. We used a

very focused statistical contrast analysis that indicated that these differences were consistent with our hypothesis, which suggests that group entrainment is modulated by a combination of expertise and familiarity with the partner. However, the analysis only compares the combined hypothesis to the data, and not the individual factors (expertise & matchedness of pairs) directly. Therefore, although matchedness had more weight in our model than expertise, we can not based on this analysis conclude anything about the relative contributions of these two factors. Also, this test does not rule out the possibility that the differences are due to some other, unobserved variable. Therefore the study needs to be expanded and control conditions added to eliminate confounding variables. It would be preferable to replicate this study with another combination of cultural and musical backgrounds.

8.4 General discussion

We conducted two motion capture experiments in African music and dance. In both these studies, the settings were cross-cultural, but the differences in expertise and familiarity seemed to us to be the more salient factors distinguishing the two groups.

These studies have contributed to two themes in music cognition research: embodiment of metre and group synchronisation. The comparative setups allowed us to test and demonstrate the potential of our analysis methods, as well as provided us with very interesting descriptive data.

The embodied nature of music cognition was discussed in chapter 2. Previous research (e.g. Phillips-Silver & Trainor, 2005, 2007) has shown that bodily movements have a central role in our understanding of rhythm and metre, guiding our perception especially when the rhythm is ambiguous. Another example of this is the study on the Easter songs of the Northern Potosi in Bolivia (Stobart & Cross, 2000). From a Western perspective, it is easy to mistakenly perceive these songs as being in a compound $\frac{6}{8}$ time starting with an anacrusis, while observing the local dancers revealed that their footfalls actually mark it as $\frac{2}{4}$ starting with a downbeat. A similar difference, although a simpler sixteenth-note phase shift was observed in cross-cultural tapping studies by Toivainen and Eerola (2003). Their South African participants perceived simplified MIDI renditions of some South African folk melodies as being continuously syncopated, while their Finnish counterparts

perceived the notes to fall on the beats rather than just after them.

In our studies, novices and experts differed from each other in how they conveyed the different metrical levels of the music in their body movements. In the Dance Study, novices seemed to entrain at a lower metrical level than the experts (twice as fast). This could be a reflection of expertise; Drake et al. (2000) showed that adults and participants with musical training were able to synchronise at a wider range of tempi, and they also spontaneously focused their attention to a higher metrical level (slower) than did children and non-musicians.

In the Choir Study, the movement trajectory of the expert's head marker suggested multiple metrical levels. This result is related to what Toivainen et al. (2010) found in spontaneous movement to music: body movements can be broken down to their principle components, each component representing one metrical level, and by animating these movement components (eigenmovements), it is easy to see that different body parts are responsible for different metrical levels. The head movement would seem to capture them all, however, and in our case the faster beat-level motion was accompanied with a slower wave at the bar-level. Meanwhile, our novices seemed to focus all their attention to the tactus and their feet. Their steps took place at the beat-level, as did the novice's whole body, as she had to pay a lot of attention to remembering the movements, the melody and the lyrics, and there was little left for expression or enactment of other metrical levels.

This result makes sense not only from the perspective of attentional resource theories (Keller, 1999), but also from the point of view of motor programs and performance studies. Brain imaging studies have shown that the physical familiarity of movements modulates the activation of the action-observation network in the brain: if you know how the movement is done, the activation is increased (Calvo-Merino et al., 2005, 2006; E. S. Cross, de C. Hamilton, & Grafton, 2006). This increased activation likely reflects the more fine-tuned motor programs one has developed and trained for executing those movements, and they might also reflect the deeper understanding of the intentions behind them (Davidson & Good, 2002; Palmer, 1989, 1997; Wanderley et al., 2005).

One additional aspect of the Choir Study worth considering is the learning process. Having new musical material to learn in an unfamiliar style

gives an excellent window into human learning processes. One of the findings of recent studies indicates that learning by observing activates the action-observation network just as learning by doing does (Grafton & Cross, 2008). The combination of both doing and observing seems to be the most effective method for learning new motor skills, such as dance movements, perhaps because one can build both internal and external views to the motor activities in parallel. Also, learning in group has been shown to accelerate learning (Mattar & Gribble, 2005), probably partly through increased motivation that is often fueled by the emotional and social benefits that come even from simple and brief experiences of successful entrainment (Hove & Risen, 2009). In the Choir Study, the response from both groups was overwhelmingly positive; they had fun in the workshop and felt motivated and satisfied with the experience. The novices learned new songs in a new style very quickly, even though they also needed to learn them aurally, which is not the way they are used to when learning new songs.

In general, we found that motion capture is an excellent tool for studying musical interaction and entrainment. The markers are unobtrusive and even in the workshop situation, participants are quick to forget they even wear them. In our setup, the cameras are positioned on beams that are almost at ceiling height on the walls. This probably makes it easy to forget the cameras, easier than it would be having to perform to a video camera that would need to be positioned at face level so that the resulting video image would have a proper viewing angle. As the viewing angle of the motion capture data can be chosen post hoc and at will, the cameras can be placed out of sight.

These two data sets have also provided us with a good testbed to develop methods for measuring pairwise and group synchrony. We have especially been interested in shifting away from the egocentric to the interactionist view, and thus focus the analysis on the global interaction itself rather than focusing on individual performance measures. One important step is to uncover the temporal evolution of entrainment, and the windowed cross-correlation is a good tool to do this, at least for dyads; for groups, the order parameter provides a similarly useful, continuous view of the entrainment. Extending from dyads to groups potentially complicates things. The dyadic patterns of influence for example, are simple compared to the hierarchical networks that might operate in an ensemble (e.g. Maduelli & Wing,

2007). However, a choir provides a natural setting where a larger group, e.g. sopranos or tenors, can have similar roles, which simplifies things, and influence and entrainment can be investigated between groups, as in (e.g. Lucas et al., 2011). This flexibility and range of different group dynamics make music and dance interesting contexts for the study of the mind, and helps balancing ecological validity with internal and external validity in experiments.

Chapter 9

Conclusions

9.1 The aims of the thesis

In this thesis, I have discussed an embodied, situated, and social approaches to studying music cognition, in the context of timing and entrainment in musical interaction. I have outlined the theoretical background for this approach, drawing from research in various fields, mainly outside music psychology. This is partly because in music cognition research, the point of view is largely a “passive” perception of music, and when the production of music is considered, it is often modelled on Western, classical music (I. Cross, 2012). In general, *participatory* musical practices are more poorly understood from cognitive point of view than *presentational* ones (Turino, 2008). There are many exceptions and the field is changing fast and adopting a more inclusive definition of “music”; I hope this thesis contributes to accelerating this process.

I have argued, along the lines of Richardson et al. (2010), that in order to understand how the mind works, we must shift from the “egocentric” model of the mind to an “interactionist” one. In making this shift, measurement and analysis methods play a central role. For this reason, much of my thesis project has been about adopting, developing and applying methods to analyse interaction, especially the temporal interrelationships in dyads and groups.

The two empirical studies presented in this thesis were preceded by a number of pilot studies, various smaller tests and trials. These pilot studies are briefly summarised in appendix A. These studies have since been followed

by more advanced and more elegant experiments, which I will briefly discuss at the end of this chapter. The two empirical studies in chapters 7 and 8 are examples of the interactionist approach, and they represent extensions of the traditional cognitive approach. The cooperative tapping study adds a social dimension, and the African dance study adds an embodied dimension. In the next sections, I will discuss the main results of these experiments and how they tie in with the theoretical background. Their results are somewhat interesting in themselves, but I see them mostly as “proof of concept studies” where the interactionist, social, and embodied concepts were put into practice, and analysis methods were tested and developed.

My initial idea was to conduct cross-cultural studies on music cognition. Based on my master’s thesis (Eerola, Himberg, Toivainen, & Louhivuori, 2006; Himberg, 2002) and all that had been written about differences and similarities between “Western” and “African” concepts of rhythm (Agawu, 1995; Temperley, 2000), rhythm seemed like a domain of music cognition that would be especially interesting from a cross-cultural point of view. I was also inspired by findings of *cultural psychology* (Cole, 1996), and Cole’s studies where he showed that people with different cultural backgrounds do well in different kinds of (intelligence) tests. An early idea was to construct a battery of rhythmic production tasks that would test various aspects of rhythmic skills, including interactions between participants. I felt that it was important to include the social aspect of musicking in the setting, as that is such an important component in many music cultures. I hypothesised, to use Thomas Turino’s terms, that growing up in a *participatory* musical culture would equip a person with a different set of rhythmic and coordination skills than learning a *presentational* musical practice, such as Western classical music (Turino, 2008).

My intention was to select a set of interactive, dyadic rhythm tasks and a set of methods from the extensive rhythm production literature, and use them to build a cross-cultural experiment. Surprisingly, no such tasks or methods existed. The rhythm production literature was based on individual production tasks. Thus, my project pivoted into developing those interactive tasks and methods. The guiding principle from the beginning was to come up with something that would be flexible and naturalistic enough to be used in a cross-cultural context (Berry et al., 1997). Later, as more books and papers on embodied, situated and social cognition started to appear, at the

same time as I was conducting my research, the theoretical framework of the project started to take shape. I feel that in the African choir and dance studies I achieved my aim to conduct cross-cultural research on embodied social cognition. These studies focused on interaction of the participants rather than their mind/body states, just as the interactionist approach proposes (M. Richardson, Marsh, & Schmidt, 2010).

9.2 Embodiment

Music is a complex physical and mental activity, and especially good in illustrating the integral role that our bodies play in the workings of our minds. Embodied and enactive approaches to cognition acknowledge this connection of the mind and the body; embodied music cognition (e.g. Leman, 2008) is about the mechanisms and processes that link musical thought to movements and sounds. The body communicates musical ideas, and in turn, shapes those ideas. The body can serve as a memory as well as a metronome; the tactus is “banked” to the taps of the foot or to the nods of the head, as larger body parts help in keeping the tempo while the finger muscles execute the finely controlled motor commands at a millisecond accuracy. The results of the African dance experiment added supporting evidence to the idea that metre is embodied. Different parts of the body move to different metrical levels, and even complicated rhythms start to “make sense” when you dance to them.

Recent rapid development of motion capture and movement analysis methods have enabled empirical study of the embodied approach. Consequently, there have been many definitions of what embodiment means, and embodied cognition comes in many flavours. Many proponents of the embodied view of cognition, for example A. D. Wilson and Golonka (2013), emphasise the dynamic nature of embodiment, and that in embodied cognition, no mental representations are needed. In my view, defining what is included in the cognitive system is crucial. In the embodied approach it consists of the mind-in-the-body and the environment, rather than just the individual mind.

Musical sounds are (usually) produced by body movements. Also, musical sounds are sounds “of something” (E. F. Clarke, 2005). Musical meaning comes from many sources, but part of the “aboutness” of music is derived

from the evolutionarily ancient motivational-structural sources (I. Cross, 2005, 2009), and universal “psychophysical cues” that are shared across musical cultures (Balkwill & Thompson, 1999). Music is born from movement, and as a temporal phenomenon it is also described using movement-related and spatial terms. We speed up, slow down, have high and low notes, and end musical pieces by slowing them down as if slowing down from movement (Friberg & Sundberg, 1999; Honing, 2003). Music and dance go hand in hand, more so in some cultures than in others (Blacking, 1973), but from an embodied cognitive point of view, making sharp distinctions between music and dance is not necessary nor helpful. In our African dance study, the choir members would tie shakers around their ankles, melting together the dynamics of their movements and those of the musical sounds. Figuring out what is going on cognitively would be impossible using the framework of traditional cognitive science with its rigid definition of discrete cognitive domains. This example illustrates that it is important to extend the cognitive system by including the social environment integrally into the model of cognition. Just as with the body, it is not enough to just acknowledge its presence by allowing it to influence the otherwise “insulated” cognition.

9.3 Social cognition

In traditional cognitive research participants are tested individually, under very controlled circumstances. This has many advantages, for example having full control over all the stimuli for each participant, contexts are identical for all participants, their performance can be compared to objective baselines and so forth (being in perfect synchrony with the metronome, or achieving perfect isochrony in solo tapping can be considered such baselines, endpoints of measuring scales for performance). And in these solo tasks, they also seem to be meaningful targets for the participants.

In contrast, when two people are interacting with each other, one person serves as the stimulus for the other, and vice versa. Their relationship, and with it the circumstances of the experiment, are in constant flux. While perfect synchrony or perfect isochrony might still be pragmatic as a baseline measure or an endpoint of a scale, it is important to remember that for the participants they might no longer represent meaningful goals. Interaction engages both participants to such an extent that keeping to-

gether becomes the primary goal over maintaining perfect isochrony, and the constantly evolving asynchrony between participants (mutual adaptation) is much more “organic” than the “mechanistic” relationship between solo taps and the metronome beats. In cooperative tapping experiments performed for this thesis it was repeatedly observed that the engagement between participants would easily override the original objective of keeping a constant tempo.

What makes interpersonal entrainment so engaging? Perhaps it is the *mutuality* of adaptation—the ongoing process of taking each other into account—through which synchrony is achieved. While this mechanism is something we seem to use automatically, it also provides us socially relevant information about the partner and the interaction. Mutual rhythmic adaptation is closely related to subconscious mimicry of gestures and convergence of speech rhythms between conversation partners, and like most forms of behavioural coordination, it is a source of positive affect. In tapping studies, with discrete data (the oscillatory process is sampled only at beat onset times), the signature of mutual adaptation is that both participants seem to follow the other simultaneously, in a patterns of +1/-1 lag positive autocorrelation (see e.g. Konvalinka et al., 2011). In our African dance experiment, continuous movement allowed us to visualise the evolution of entrainment at a finer resolution. This made it quite clear that what in the static view looks like zero-lag synchrony, is actually a product of constantly evolving mutual adaptation.

In the computer tapper experiment, the difference between one-sided and mutual adaptation was demonstrated, when the human-human pair scored higher on the entrainment measure than human-metronome pairs, even though the metronome helped to maintain a steady tempo. The human pairs drifted in tempo, but locked their relative phase. The goal of synchrony was achieved better than with a perfectly predictable and stable metronome. This result was surprising and strongly supports the idea that mechanisms of social cognition such as rhythmic entrainment need to be studied in actual social settings.

9.4 Shared manifold - a model for social cognition

Vittorio Gallese, a member of the research team in Parma that originally discovered mirror neurons, has attempted to contextualise the 20 year old finding and relate it to cognitive processes (Gallese, 2001, 2003, 2005). He writes about “shared manifold”, referring to a number of interpersonal processes that are based on the operations of mirroring systems, and facilitate interpersonal understanding and empathy (see figure 9.1). His model is an attempt to bring together different aspects of sociality and embodiment, and the results obtained in this thesis also support this model.

The “manifold” refers to the idea that essentially the same system is responsible for sharing many kinds of things, also at many different timescales. Gallese sees that simulation is the mechanism behind all kinds of interpersonal interaction. We understand others by virtue of them being “like us”. This idea is familiar from the theory-theory, or the theory of mind, but is more embodied and assumes fewer cognitive constructions and representations. In effect, instead of taking sides with either the “cognitive constructionists” or the “radically embodied” crowd, for Gallese imitation, empathy, and mind-reading all share a fundamental feature: they all need a “shared, meaningful intersubjective space” (Gallese, 2003, 517).

For Gallese, intersubjectivity means sharing actions, emotions, intentions, sensations and well as body schemata, as listed in the blue and green bubbles representing “Agents” in figure 9.1. This sharing is central in any joint action; collaboration and coordination are based on this mutual understanding of goals and actions. Humans have a high motivation for sharing goals and intentions, as part of our natural sociality (Hrdy, 2009; Tomasello & Carpenter, 2007; Tomasello et al., 2005). Music and dance are prime examples of such joint actions, due to the way they link emotion and action to complex social collaboration.

Gallese envisions this sharing and intersubjectivity as taking place at three different levels (red bubble in figure 9.1). At the bottom, the sub-personal level (see the dashed box), there are the neural mechanisms that underpin these processes. These include the mirror neuron system, as well as the limbic system and the various neurotransmitter and hormone related processes producing the various body states that go with our emotions. These processes are usually studied by neuroscientists and while much of

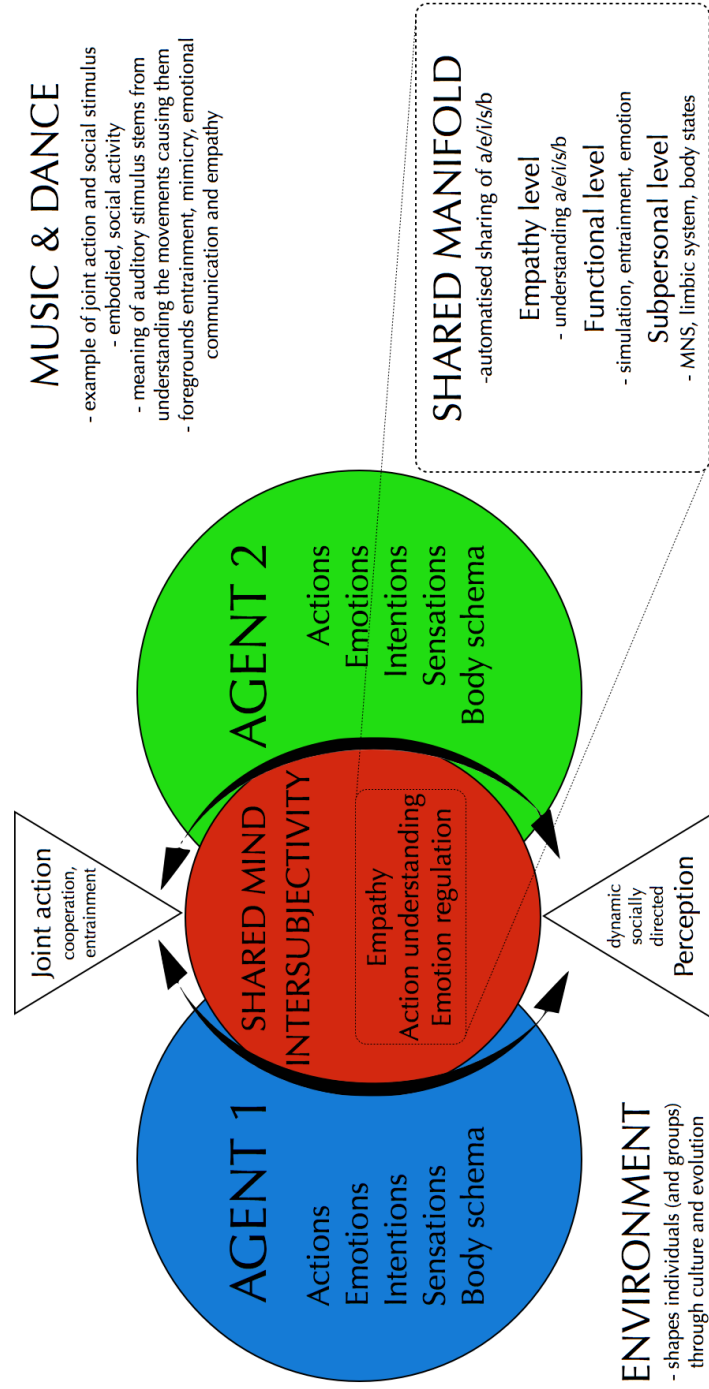


Figure 9.1: Shared manifold - Gallese's framework for intersubjectivity.

this research is done in individuals (at the subpersonal level), new, social approaches are emerging (Hari & Kujala, 2009).

At the functional level (dashed box), we can find basic processes of joint actions: entrainment, action simulation and emotional contagion. This is the level at which most social cognition research is done. These are processes that are subserved by the mechanisms of the subpersonal level. They can and have been experimented on as such, but it is important to see how they connect to the other levels, as well as how they link to the environment. For example, studying entrainment is a way to explore the “shared space”. Synchronisation is an extreme form of imitation; being temporally simultaneous, a kind of real-time predictive simulation of the actions of the others.

At the top level (dashed box) we find understanding of others, or the empathetic level. This is the phenomenological level of interpersonal understanding and connection that arises from the functionality of entrainment and emotional contagion, based on the fundamental way we are wired to share our intentionality.

Gallese’s model (excellently discussed by Semin & Smith, 2008), can also easily be connected with what we know about joint action from the social cognition literature (Knoblich & Sebanz, 2008; Sebanz, Bekkering, & Knoblich, 2006), and from the dynamic approaches to interaction and interpersonal entrainment (M. Richardson et al., 2007; M. J. Richardson, Marsh, & Schmidt, 2005; R. Schmidt & Richardson, 2008; Shockley, Richardson, & Dale, 2009; Shockley et al., 2003). Also, the field of ecological and dynamic perception and attention is compatible with the levels of intersubjectivity (Carpenter et al., 1998; E. F. Clarke, 2005; Keller, 1999; Langton, Watt, & Bruce, 2000; Large & Jones, 1999).

Entrainment is one of the key processes in Gallese’s model. We know from a wide range of research that it is automatic, and connected to emotions (e.g. Hove & Risen, 2009; Zentner & Eerola, 2010), just like Gallese’s model predicts. The links between emotions, intentions, actions and synchronisation of brain activities have recently been studied using fMRI (Nummenmaa et al., 2012). In this imaging study, participants watched emotion-eliciting film segments, and these shared emotional experiences were found to synchronise their brain activations in an emotion-specific fashion, especially in those brain regions that are central in understanding others’ intentions and actions. However, given the approximately 1-3 second temporal resolution

of fMRI, this “synchronisation” is at a different functional level from the millisecond-level, phase-locking, fast synchronisation measured in dual EEG or MEG studies (e.g. Babiloni & Astolfi, 2012; Baess et al., 2012; Sanger et al., 2011, 2012; Tognoli, Lagarde, DeGuzman, & Kelso, 2007; Yun et al., 2012). Also, as Nummenmaa et al. scanned people one at a time, their findings are based on shared temporal trajectories of activation patterns rather than actual real-time entrainment, but they are illuminating nevertheless.

Mutual adaptation, the automatic, two-sided process of keeping a dyad in time and in sync was observed in all the experiments reported in this thesis, regardless of whether participants were tapping their fingers or dancing in groups or in dyads, and whether they were experts or novices, musicians or non-musicians. This supports the idea that entrainment is an automatic, basic level process that supports the higher-level social functions. Through facilitating action understanding, entrainment serves in constructing meaning in social interaction. For example, metrical structures are important building blocks of music, and guide the perception of musical ideas (Lerdahl & Jackendoff, 1983; London, 2012; London, Himberg, & Cross, 2009; Stobart & Cross, 2000). The metrical organisation of the beat is embodied, in that different body parts enact different levels of the metrical hierarchy (Toiviainen et al., 2010). Our African dance studies demonstrated how different levels of familiarity and different musical backgrounds have an effect on patterns of entrainment; shared meaning corresponds with and arises from shared movement patterns.

Embodiment of metre is one example of the way we understand music, and how music intentionalises time. Musicking bodies move in predictable patterns, and given, for example, a shared concept of the “origin” of these movements (hips in the case of our South African dances), movement patterns communicate these concepts to others, and from the perception of mutuality in movement, a positive social and emotional response is elicited. As all humans share the general composition of their bodies, such a level of even universal, mutual understanding is possible. However, the culturally contingent ways of moving in time can add layers upon layers of meanings that are only available for those who share that cultural background.

While the more “universal” features of music afford meaningful interactions between people from different musical backgrounds, cultural differences do matter, and certain levels of meaning that experts can access, are not

available for novices. Collaboration between ethnomusicologists and music cognition experts can be very fruitful in exploring the fundamentals of music perception and processing (e.g. Eerola, Louhivuori, & Lebaka, 2009; Fritz et al., 2009; Krumhansl, Louhivuori, Toiviainen, Jarvinen, & Eerola, 1999; Krumhansl et al., 2000). In the light of our findings, it would be especially important to adopt the embodied, social approach to cognition in these studies. This would help to ensure that a wide enough range of musical behaviours is considered and that the focus is not exclusively on symbolic information processing.

9.5 Future directions

The research carried out for this thesis has generated many ideas for further research. Many of these studies are already ongoing, and some results have been published. They are all closely linked to the theme of this thesis, and most utilise methods discussed here (see appendix E).

To explore whether musical and verbal communication *share* similar entrainment mechanisms, we designed a study to compare patterns of entrainment in dyads engaged in musical and verbal improvisations (Gill et al., 2012). In this study, we are combining qualitative and quantitative analyses of interaction. The qualitative analysis (based on Gill, 2007) steers the quantitative analysis by flagging those segments where the most meaningful and important interactions take place. Cross-recurrence analysis (see chapter 5.2) is then employed to uncover patterns of entrainment, and kinematic analysis is used to describe how bodies manifest these pivotal moments in the interactions.

This study follows a recent interest in studying the nexus of music and language. For example, both music and language can be seen as symbol-based information processing systems and thus similar tools and concepts can be used in analysis and modelling, e.g. (Lerdahl & Jackendoff, 1983). The more recent research is more interested in the shared neural underpinnings of language and music (Koelsch, 2009; Patel, 2010; Sammler et al., 2012, e.g.), or are looking at dialogue as joint action (Garrod & Pickering, 2004, 2009), or try to understand the basis of communication systems in general by looking at “novel forms of communication which people develop when they cannot use pre- established communication systems”(Galantucci & Garrod,

2011, 1), a field called *experimental semiotics* (Galantucci, 2009; Galantucci & Garrod, 2011).

Only relatively few researchers so far have been interested in entrainment in conversation, even though interpersonal timing plays a very important role in conversational turn-taking (Cummins, 2005, 2009; M. Wilson & Wilson, 2005). Also, in studying conversation as joint action, temporal alignment and convergence of speech rate and synchronisation of gestures serve as important cues for gauging the success of the communication (Gill, 2007). Cross-pollination between the language and music research communities is already showing to be very fruitful, and this interdisciplinary work is very useful for all involved.

The links between entrainment and the emotional consequences are an important but as of yet, under-researched area. Music, the entrainer-engager *par excellence*, has the capability of inducing strong emotions (Gabrielsson, 2001), and it is also efficacious in emotion regulation and in music therapy interventions related to emotional processes (Kim, Wigram, & Gold, 2009; Saarikallio & Erkkilä, 2007; Thaut et al., 2009). Group music-making improves trust and cooperation (Anshel & Kipper, 1988) as well as general psychological well-being (Clift et al., 2007, 2010; Cohen et al., 2006), and reduces stress (Kreutz et al., 2004). The mechanisms for these effects are not yet well-known, and links between empathy and entrainment, for example, are only explored in a few studies (Himberg & Spiro, 2012; Hove & Risen, 2009; Rabinowitch, Cross, & Burnard, 2012; Spiro & Himberg, 2012). Our project aims to explore these mechanisms systematically (e.g. using resistance to entrainment as a task (Spiro & Himberg, 2012)). A related project will also look at patterns of interaction in music therapy improvisations, and explore novel ways of analysing them. Preliminary findings from various pilot studies on entrainment and autism (see e.g. chapter 6) suggest that the automatic mutuality of entrainment might be working differently in individuals within the autism spectrum. While there is a growing body of research on autism and music therapy, much of this research is qualitative. This research of course helps practitioners understand the phenomenon and reflect on the therapeutic process, but it does not typically drill down to the mechanisms of how the chosen music therapy works, or even what the underlying mechanisms are. Our project aims to bring together the qualitative understanding of music therapists with the quantitative methodology

of cognitive science.

There is a trend towards more naturalistic research settings also in neuroscience. Although the methods are still quite restrictive, there is also increased interest in studying embodied and social cognition. Two-person neuroscience (Hari & Kujala, 2009) is quickly picking up pace. Through developing better equipment and more robust acquisition and analysis methods, more naturalistic stimuli and tasks can be used. In my current lab, a two-person MEG setup has been developed (Baess et al., 2012), as well as an implementation of dual fMRI (Renvall & Malinen, 2012). The MEG2MEG setup connects two MEG sites together so that two people can interact with each other, and their brain activity can be recorded and synchronised. There are two-person setups where only one participant is scanned or recorded at the time, but given the dynamic nature of interactions and the temporal scale of mutual adaptations, it is necessary to be able to measure the activity of the “hyperbrain”, the combined network of the two brains. The extended view of cognition that Hutchins (1995) has promoted is no longer just a metaphor, but the interacting brains seem to be genuinely entrained and working “as one”. Again, $1 + 1$ seems to be more than 2.

In two-person EEG and MEG studies, various brain connectivity, network analysis, and synchronisation measures have been developed (Lindenberger et al., 2009; Sanger et al., 2012; Yun et al., 2012). The measures used to map networks within individual brains can be extended for studying the “hyperbrain”. This line of studies is still so new that there are no canonical methods or measures, even the philosophical basis of these studies is underdeveloped. However, the envelope is constantly being pushed in terms of behavioural tasks, measurements, as well as analysis methods and visualisations. The limitations and restrictions are being circumvented one by one. Combined with an increasing awareness of the need for more naturalistic settings and the rapid development of the equipment and research methods this is going to be a very interesting field in the coming years. I hope that the theories, methods and empirical findings presented in this thesis will in some small way help form the basis of this future research.

References

- Acebrón, J., Bonilla, L., Vicente, C., Ritort, F., & Spigler, R. (2005). The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1), 137–185.
- Agawu, K. (1995). The invention of “african rhythm”. *Journal of the American Musicological Society*, 380–395.
- Allport, F. H. (1924 (reprint 1994)). *Social psychology* (reprint 1994 ed.). Routledge/Thoemmes Press.
- Anderson, T. (1971). *The statistical analysis of time series* (R. A. Bradley, J. S. Hunter, D. G. Kendall, & G. S. Watson, Eds.). New York: John Wiley & Sons, Inc.
- Anshel, A., & Kipper, D. (1988). The influence of group singing on trust and cooperation. *Journal of Music Therapy*, 25(3), 145–155.
- Arrindell, W. A., & van der Ende, J. (1985). Cross-sample invariance of the structure of self-reported distress and difficulty in assertiveness : Experiences with the scale for interpersonal behaviour experiences with the scale for interpersonal behaviour. *Advances in Behaviour Research and Therapy*, 7(4), 205–243.
- Aschersleben, G., & Prinz, W. (1995, 1995/4). Synchronizing actions with events: the role of sensory information. *Perception & Psychophysics*, 57(3), 305-317.
- Auslander, P. (2008). *Liveness: Performance in a mediatized culture* (2nd ed.). Abingdon, UK: Routledge.
- Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New ideas in psychology*, 27(1), 32–47. doi: 10.1016/j.newideapsych.2007.12.002
- Babiloni, F., & Astolfi, L. (2012). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience & Biobehavioral*

- Reviews, in press.* doi: 10.1016/j.neubiorev.2012.07.006
- Baess, P., Zhdanov, A., Mandel, A., Parkkonen, L., Hirvenkari, L., Mäkelä, J., ... Hari, R. (2012). Meg dual scanning: a procedure to study real-time auditory interaction between two persons. *Frontiers in Human Neuroscience*, 6.
- Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, 17(1), 43-64.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, 22(04), 577-660.
- Bavelas, J., Black, A., Lemery, C., & Mullett, J. (1986). I show how you feel: Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, 50(2), 322.
- Beck, R. J., Cesario, T. C., Yousefi, A., & Enamoto, H. (2000). Choral singing, performance perception, and immune system changes in salivary immunoglobulin a and cortisol. *Music Perception*, 18(2), 87-106.
- Berens, P. (2009). Circstat: A matlab toolbox for circular statistics. *Journal of Statistical Software*, 31(10), 1-21.
- Bernieri, F. J., Davis, J. M., Rosenthal, R., & Knee, C. R. (1994). Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin*, 20(3), 303-311. doi: 10.1177/0146167294203008
- Bernieri, F. J., Reznick, J., & Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology*, 54(2), 243-253. doi: 10.1037/0022-3514.54.2.243
- Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior*. Cambridge: Cambridge University Press.
- Berry, J., Poortinga, Y., & Pandey, J. (1997). *Handbook of cross-cultural psychology: Basic processes and human development* (Vol. 2). Boston, MA: Allyn & Bacon.
- Bispham, J. (2006). Rhythm in music: What is it? who has it? and why?

- Music Perception*, 24(2), 125–134. doi: MP.2006.24.2.125
- Blacking, J. (1973). *How musical is man?* Seattle: University of Washington Press.
- Boker, S. M., Rotondo, J. L., Xu, M., & King, K. (2002, Sep). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, 7(3), 338–355.
- Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528), 2359. doi: 10.1098/rstb.2009.0066
- Braun, D. A., Ortega, P. A., & Wolpert, D. M. (2011). Motor coordination: when two have to act as one. *Experimental brain research*, 211(3–4), 631–641. doi: 10.1007/s00221-011-2642-y
- Brown, J. C. (1993). Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94(4), 1953–1957.
- Brownlow, S., & Dixon, A. R. (1997). Perception of movement and dancer characteristics from point-light displays of dance. *Psychological Record*, 47(3), 411.
- Bruner, J. S. (1964). The course of cognitive growth. *American psychologist*, 19(1), 1–15.
- Bruscia, K. (1987). *Improvisational models of music therapy*. CC Thomas, Springfield, Ill., USA.
- Buck, J. B. (1938). Synchronous rhythmic flashing of fireflies. *The Quarterly Review of Biology*, 13(3), 301–314.
- Buck, J. B., & Buck, E. (1968). Mechanism of rhythmic synchronous flashing of fireflies: Fireflies of southeast asia may use anticipatory time-measuring in synchronizing their flashing. *Science*, 159(3821), 1319–1327. doi: 10.1126/science.159.3821.1319
- Burgoon, J. K., Stern, L. A., & Dillman, L. (Eds.). (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge, UK: Cambridge University Press.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365–376. doi: 10.1038/nrn3475

- Cacioppo, J., Berntson, G., Sheridan, J., & McClintock, M. (2000). Multi-level integrative analyses of human behavior: social neuroscience and the complementing nature of social and biological approaches. *Psychological bulletin*, 126(6), 829-843.
- Calvo-Merino, B., Glaser, D., Grèzes, J., Passingham, R., & Haggard, P. (2005). Action observation and acquired motor skills: an fmri study with expert dancers. *Cerebral cortex*, 15(8), 1243-1249.
- Calvo-Merino, B., Grèzes, J., Glaser, D., Passingham, R., & Haggard, P. (2006). Seeing or doing? influence of visual and motor familiarity in action observation. *Current Biology*, 16(19), 1905-1910.
- Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, 110(1-2), 43 - 50. doi: DOI: 10.1016/S0167-2789(97)00118-8
- Cappella, J. N. (1991). Mutual adaptation and relativity of measurement. In B. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (Vol. 1, pp. 103-117). New York: Guilford Press.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs Of The Society For Research In Child Development*, 63(4), i-vi, 1-143.
- Carson, R. G., Goodman, D., Kelso, J. A. S., & Elliott, D. (1995). Phase transitions and critical fluctuations in rhythmic coordination of ipsilateral hand and foot. *Journal of Motor Behavior*, 27(3), 211-224. doi: 10.1080/00222895.1995.9941711
- Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interactions. *Journal of personality and social psychology*, 76, 893-910.
- Chatfield, C. (2004). *The analysis of time series: An introduction* (6th edition ed.). Boca Raton, FL: CRC Press.
- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
- Clarke, E. (1982). Timing in the performance of erik satie's 'vexations'. *Acta Psychologica*, 50(1), 1-19. doi: 10.1016/0001-6918(82)90047-6
- Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford: Oxford University Press.
- Clayton, A. (1986). *Coordination between players in musical performance*.

- Unpublished doctoral dissertation, Edinburgh University, Edinburgh, UK.
- Clayton, M., Sager, R., & Will, U. (2004). In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM CounterPoint*, 1, 1-45.
- Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2007). Choral singing and psychological wellbeing: Findings from english choirs in a cross-national survey using the whoqol-bref. In *Proceedings, international symposium on performance science, porto, portugal* (pp. 22–23).
- Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2010). Choral singing and psychological wellbeing: Quantitative and qualitative findings from english choirs in a cross-national survey. *Journal of Applied Arts and Health*, 1(1), 19–34. doi: 10.1386/jaah.1.1.19/1
- Cohen, G., Perlstein, S., Chapline, J., Kelly, J., Firth, K., & Simmens, S. (2006). The impact of professionally conducted cultural programs on the physical health, mental health, and social functioning of older adults. *The Gerontologist*, 46(6), 726.
- Cole, M. (1996). *Cultural psychology – a once and future discipline*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Condon, W., & Ogston, W. (1967). A segmentation of behavior. *Journal of psychiatric research*, 5, 221-235.
- Coupland, J., Coupland, N., & Robinson, J. D. (1992). “how are you?”: Negotiating phatic communion. *Language in Society*, 21, 207–230.
- Cross, E. S., de C. Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, 31(3), 1257 - 1267. doi: 10.1016/j.neuroimage.2006.01.033
- Cross, I. (2001). Music, cognition, culture and evolution. *Annals of the New York Academy of Sciences*, 930, 28-42.
- Cross, I. (2005). Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 27–43). Oxford: Oxford University Press.
- Cross, I. (2008). Musicality and the human capacity for culture. *Musicae Scientiae*, 12(1), 147–167.
- Cross, I. (2009). The nature of music and its evolution. In S. Hallam, I. Cross,

- & M. H. Thaut (Eds.), *The oxford handbook of music psychology* (pp. 3–13). Oxford: Oxford University Press.
- Cross, I. (2012). Cognitive science and the cultural nature of music. *Topics in Cognitive Science*, 4(4), 668–677. doi: 10.1111/j.1756-8765.2012.01216.x
- Cross, I., Fitch, W. T., Aboitiz, F., Iriki, A., Jarvis, E. D., Lewis, J., ... Trehub, S. E. (2013). Culture and evolution. In M. A. Arbib (Ed.), *Language, music, and the brain* (Vol. 10, pp. 541–562). Cambridge, MA: MIT Press.
- Cummins, F. (2005). Interval timing in spoken lists of words. *Music Perception*, 22(3), 497–508. (Journal Article)
- Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37, 16–28. doi: doi:10.1016/j.wocn.2008.08.003
- Davidson, J. W. (2009). Movement and collaboration in musical performance. In S. Hallam, I. Cross, & M. H. Thaut (Eds.), *The oxford handbook of music psychology* (pp. 364–376). Oxford University Press.
- Davidson, J. W., & Correia, J. S. (2002). Body movement. In R. Parncutt & G. McPherson (Eds.), *The science and psychology of music performance* (pp. 237–252). New York, NY: Oxford University Press.
- Davidson, J. W., & Good, J. (2002). Social and musical co-ordination between members of a string quartet: An exploratory study. *Psychology of Music*, 30(2), 186–201.
- Davis, J. I., Benforado, A., Esrock, E., Turner, A., Dalton, R. C., van Noorden, L., & Leman, M. (2012). Four applications of embodied cognition. *Topics in Cognitive Science*, 4(4), 786–793. doi: 10.1111/j.1756-8765.2012.01225.x
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85–103.
- De Bruyn, L., Leman, M., Moelants, D., & Demey, M. (2009). Does social interaction activate music listeners? In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music, modeling and retrieval. genesis of meaning of sound and music. lecture notes in computer science* (pp. 93–106). Berlin: Springer-Verlag.
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phe-*

- nomenology and the Cognitive Sciences*, 6, 485-507. (10.1007/s11097-007-9076-9) doi: 10.1007/s11097-007-9076-9
- Desain, P., & Honing, H. (1999). Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, 28(1), 29–42.
- Deutsch, D. (1983). The generation of two isochronous sequences in parallel. *Attention, Perception, & Psychophysics*, 34(4), 331–337.
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychological review*, 3(4), 357. doi: 10.1037/h0070405
- Drake, C., Jones, M. R., & Baruch, C. (2000, 2000/12/15). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, 77(3), 251-288. (TY - JOUR)
- Dunlap, K. (1910). Reactions to rhythmic stimuli with attempt to synchronize. *Psychological Review*, 17, 399–416.
- Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, 4(9), 973-977.
- Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and african folk melodies by western and african listeners. *Psychology of Music*, 34(3), 337-371.
- Eerola, T., Louhivuori, J., & Lebaka, E. (2009). Expectancy in sami yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicae Scientiae*, 13(2), 231–272.
- Eerola, T., Luck, G., & Toiviainen, P. (2006). An investigation of pre-schoolers' corporeal synchronisation with music. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th international conference on music perception and cognition* (pp. 472–476). Bologna, Italy.
- Eerola, T., & Toiviainen, P. (2003). Midi toolbox: Matlab tools for music research (1.0 ed.) [Computer software manual]. Jyväskylä.
- Farhat, A., Amiri, R., Karbandi, S., Esmaily, H., & Mohammadzadeh, A. (2010). The effect of listening to lullaby music on physiologic response and weight gain of premature infants. *Journal of Neonatal-Perinatal Medicine*, 3(2), 103–107.
- Finney, S. A. (2001). Ftap: a linux-based program for tapping and music experiments. *Behavior Research Methods, Instruments, & Computers: a Journal Of The Psychonomic Society, Inc.*, 33(1), 65-72.

- Fisher, N. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
- Fodor, J. (1983). *The modularity of mind* (Vol. 341). Cambridge, MA: MIT press.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music*. (p. 149-180). New York:: Academic Press.
- Fraser, A. M., & Swinney, H. L. (1986, Feb). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, *33*, 1134–1140. doi: 10.1103/PhysRevA.33.1134
- Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? a model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America*, *105*(3), 1469–1484.
- Friberg, A., & Sundstrom, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, *19*(3), 333-349. (Journal Article)
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, *9*(10), 474–480. doi: 10.1016/j.tics.2005.08.011
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., ... Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, *19*(7), 573–576.
- Fuchs, A., & Jirsa, V. K. (2000). The hkb model revisited: How varying the degree of symmetry controls dynamics. *Human Movement Science*, *19*(4), 425 - 449. doi: 10.1016/S0167-9457(00)00025-7
- Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8*(4), 465–486.
- Furr, R., & Rosenthal, R. (2003). Repeated-measures contrasts for "multiple-pattern" hypotheses. *Psychological methods*, *8*(3), 275.
- Gabrielsson, A. (2001). Emotions in strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 431–449). Oxford, UK: Oxford University Press.
- Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of music*, *31*(3), 221–272.
- Galantucci, B. (2009). Experimental semiotics: a new approach for studying

- communication as a form of joint action. *Topics in Cognitive Science*, 1(2), 393–410.
- Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, 5(11), 1–15. doi: 10.3389/fnhum.2011.00011
- Gallese, V. (2000). The inner sense of action. agency and motor representations. *Journal of Consciousness Studies*, 7(10), 23–40.
- Gallese, V. (2001). The shared manifold hypothesis. from mirror neurons to empathy. *Journal of Consciousness Studies*, 5(7), 33–50.
- Gallese, V. (2003). The manifold nature of interpersonal relations: A quest for a common mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358, 517–528. doi: 10.1098/rstb.2002.1234
- Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4(1), 23–48. doi: 10.1007/s11097-005-4737-z
- Garrod, S., & Pickering, M. (2004). Why is conversation so easy? *Trends in cognitive sciences*, 8(1), 8–11.
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304. doi: 10.1111/j.1756-8765.2009.01020.x
- Gibson, J. J. (1977). The concept of affordances. *Perceiving, acting, and knowing*, 67–82.
- Gibson, K. R. (1993). Tool use, language and social behavior in relationship to information processing capacities. In K. R. Gibson & T. Ingold (Eds.), *Tools, language, and cognition in human evolution* (p. 251–270). Cambridge, UK: Cambridge University Press.
- Gibson, K. R., & Ingold, T. (Eds.). (1993). *Tools, language and cognition in human evolution*. Cambridge, UK: Cambridge University Press.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press.
- Gill, S. P. (2007). Entrainment and musicality in the human system interface. *AI & Society*, 21(4), 567–605.
- Gill, S. P., Himberg, T., & Thompson, M. R. (2012). body rhythmic entrain-

- ment and pragmatics in musical and linguistic improvisation tasks. In E. Cambouropoulos, C. Tsourgas, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th international conference on music perception and cognition and the 8th triennial conference of the european society for the cognitive sciences of music* (pp. 347–348). Thessaloniki, Greece.
- Ginsborg, J., Chaffin, R., & Nicholson, G. (2006). Shared performance cues in singing and conducting: a content analysis of talk during practice. *Psychology of Music*, 34(2), 167–194.
- Goebel, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, 427–438. doi: 10.1525/MP.2009.26.5.427
- Gold, C., Wigram, T., & Elephant, C. (2006). Music therapy for autistic spectrum disorder (review). *The Cochrane Library*(4).
- Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37(1), 3–19. doi: 10.1177/0093650209351468
- Goodman, E. (2003). Ensemble performance. In J. Rink (Ed.), *Musical performance: A guide to understanding* (p. 153–167). Cambridge: Cambridge University Press.
- Grafton, S. T., & Cross, E. (2008). Dance and the brain. In C. Asbury & B. Rich (Eds.), *Learning, arts, and the brain: The dana consortium report on arts and cognition* (pp. 61–70). New York, NY: Dana Foundation.
- Grahn, J., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, 19(5), 893–906.
- Granger, C., Ghysels, E., Swanson, N., & Watson, M. (2001). *Essays in econometrics: Spectral analysis, seasonality, nonlinearity, methodology, and forecasting* (Vol. 1). Cambridge, UK: Cambridge University Press.
- Griffin, D., & Gonzalez, R. (2003). Models of dyadic social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 573–581.
- Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics*, 51(5), 347–356.

- Handel, S., & Lawson, G. (1983). The contextual nature of rhythmic interpretation. *Attention, Perception, & Psychophysics*, 34(2), 103–120.
- Hanson, F. E., Case, J. F., Buck, E., & Buck, J. (1971). Synchrony and flash entrainment in a new guinea firefly. *Science*, 174(4005), 161–164. doi: 10.1126/science.174.4005.161
- Hari, R., & Kujala, M. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiological reviews*, 89(2), 453–479.
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Paris: Cambridge University Press.
- Headland, T. N., Pike, K. L., & Harris, M. (1990). *Emics and etics: The insider/outsider debate* (Vol. 7). Newbury Park, CA: Sage Publications.
- Henry, M., & McAuley, J. (2010). On the prevalence of congenital amusia. *Music Perception*, 27(5), 413–418.
- Himberg, T. (2002). *Perception of melodic complexity - a cross-cultural investigation*. Master's thesis, University of Jyväskylä, Jyväskylä.
- Himberg, T. (2006). Co-operative tapping and collective time-keeping - differences of timing accuracy in duet performance with human or computer partner. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Icmpc 9* (p. 377). Bologna, Italy.
- Himberg, T. (2011). Interacting with non-responsive and responsive tapping partners. In *13th international rhythm perception and production workshop, poster presentation*. Leipzig, Germany.
- Himberg, T. (2012). Molemmipuolinen mukautuminen: Tahdistumisen ja musiikillisen vuorovaikutuksen tutkimus. In T. Himberg, J. Vuoskoski, & T. Eerola (Eds.), *Proceedings of the 16th annual symposium for music scholars in finland*. Jyväskylä, Finland.
- Himberg, T., & Cross, I. (2004). Interaction in musical time. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & J. P. Webster (Eds.), *Icmpc 8*. Evanston, Illinois: Causal Productions, Adelaide.
- Himberg, T., & Spiro, N. (2012). Beating to each other's drum: Towards a comprehensive understanding of pairwise rhythmic interaction. In *7th nordic music therapy congress*. Jyväskylä, Finland.
- Himberg, T., & Thompson, M. R. (2010). Dyadic entrainment and interaction in african dance. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the 11th international conference on music perception and cognition (icmpc11)*. Seattle, WA.

- Honing, H. (2003). The final ritard: on music, motion, and kinematic models. *Computer Music Journal*, 27(3), 66–72.
- Honing, H., Ladinig, O., Haden, G. P., & Winkler, I. (2009). Is beat induction innate or learned? probing emergent meter perception in adults and newborns using event-related brain potentials. *The Neurosciences and Music III—Disorders and Plasticity: Annals of the New York Academy of Sciences*, 1169, 93–96. doi: 10.1111/j.1749-6632.2009.04761.x
- Honing, H., Merchant, H., Haden, G. P., Prado, L., & Bartolo, R. (2012, 12). Rhesus monkeys (*macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE*, 7(12), e51369. doi: 10.1371/journal.pone.0051369
- Hove, M. J., & Risen, J. L. (2009). It’s all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27(6), 949–960. doi: 10.1521/soco.2009.27.6.949
- Hrdy, S. B. (2009). *Mothers and others - the evolutionary origins of mutual understanding*. Cambridge, MA: Belknap Press.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995. doi: 10.1098/rspa.1998.0193
- Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S., Laine, M., & Hämäläinen, H. (2009). Attention and cognitive control: Unfolding the dichotic listening story. *Scandinavian journal of psychology*, 50(1), 11–22.
- Huotilainen, M. (2010). Building blocks of fetal cognition: emotion and language. *Infant and Child Development*, 19(1), 94–98. doi: 10.1002/icd.658
- Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences*, 930(1), 43–61.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT press.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: The MIT Press.
- Ioannidis, J. P. A. (2005, 08). Why most published research findings are

- false. *PLoS Med*, 2(8), e124. doi: 10.1371/journal.pmed.0020124
- James, W. (1890). *The principles of psychology* (1955th ed.). Chicago: Encyclopaedia Britannica.
- Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics* (Vol. 5). Singapore: World Scientific Publishing.
- Janata, P., & Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature neuroscience*, 6(7), 682–687.
- Jensenius, A. R. (2008). *Action-sound: Developing methods and tools to study music-related body movement*. Phd thesis, University of Oslo, Oslo, Norway.
- Jensenius, A. R., Nymoen, K., Skogstad, S. A., & Voldsund, A. (2012). A study of the noise-level in two infrared marker-based motion capture systems.
- Johansson, G. (1973). Visual perception of biological motion and model for its analysis. *Perception and Psychophysics*, 14, 201–211.
- Jones, M., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological review*, 96(3), 459–491.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Keller, P. E. (1999). Attending in complex musical interactions: The adaptive dual role of meter. *Australian Journal of Psychology*, 51(3), 66–175.
- 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(1), 102–111. doi: 10.1016/j.concog.2005.12.004
- Kelley, L., Coe, R., Madden, J., & Healy, S. (2008). Vocal mimicry in songbirds. *Animal behaviour*, 76(3), 521–528.
- Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 246(6), R1000–R1004.
- Kelso, J. A. S. (1995). *Dynamic patterns - the self-organization of brain and behavior* (1st ed.). Cambridge, MA: The MIT Press.
- Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of non-

- linear, limit cycle oscillatory processes: Theory and data. *Journal of motor behavior*, 13(4), 226–261.
- Kemp, A. (1981). The personality structure of the musician: I. identifying a profile of traits for the performer. *Psychology of music*, 9(1), 3–14.
- Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta psychologica*, 32, 101–125.
- Kennel, M., Brown, R., & Abarbanel, H. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6), 3403.
- Khvedelidze, B. (2002). Hilbert transform. In M. Hazewinkel (Ed.), *Encyclopaedia of mathematics*. Springer Verlag.
- Kilner, J., Friston, K., & Frith, C. (2007). The mirror-neuron system: a bayesian perspective. *Neuroreport*, 18(6), 619–623.
- Kilner, J. M., Friston, K., & Frith, C. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, 8(3), 159–166.
- Kim, J., Wigram, T., & Gold, C. (2009). Emotional, motivational and interpersonal responsiveness of children with autism in improvisational music therapy. *Autism*, 13(4), 389–409. doi: 10.1177/1362361309105660
- King, E., & Ginsborg, J. (2011). Gestures and glances: Interactions in ensemble rehearsal. In A. Gritten & E. King (Eds.), *New perspectives on music and gesture* (pp. 177–202). Farnham, UK: Ashgate Publishing Co.
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102(3), 299–314.
- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 2021–2031. doi: 10.1098/rstb.2008.0006
- Koelsch, S. (2009). Music-syntactic processing and auditory memory – similarities and differences between eran and mmn. *Psychophysiology*, 46(1), 179–190. doi: 10.1111/j.1469-8986.2008.00752.x
- Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *The Quarterly journal of experimental psychology*, 63(11), 2220–2230. doi: 10.1080/17470218.2010.497843

- Kreutz, G., Bongard, S., Rohrman, S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin a, cortisol, and emotional state. *Journal of Behavioral Medicine*, 27(6), 623-635.
- Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Jarvinen, T., & Eerola, T. (1999). Melodic expectation in finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, 17(2), 151-195.
- Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Jarvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to north sami yoiks. *Cognition*, 76(1), 13-58.
- Kugler, P. N., & Turvey, M. T. (1987). *Information, natural laws, and self-assembly of rhythmic movement*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Kurtz, S., & Lee, T. (2003). Part and whole perceptual-motor practice of a polyrhythm. *Neuroscience letters*, 338(3), 205-208.
- LaFrance, M. (1979). Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly*, 42(1), 66-70.
- Langton, S., Watt, R., & Bruce, V. (2000). Do the eyes have it? cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2), 50-59. doi: 10.1016/S1364-6613(99)01436-9
- Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, 19(4), 527-566.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119-159.
- Large, E. W., & Kolen, J. (1994). Resonance and the perception of musical meter. *Connection science*, 6(2-3), 177-208.
- Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *International conference on digital audio effects*. Bordeaux, France.
- Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications*,. Springer Verlag.
- Leman, M. (2008). *Embodied music cognition and mediation technology*.

- Cambridge, MA: The MIT Press.
- Leman, M., Desmet, F., Styns, F., Van Noorden, L., & Moelants, D. (2009). Sharing musical expression through embodied listening: A case study based on chinese guqin music. *Music Perception*, 26(3), 263–278.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Levitin, D. J. (2007). *This is your brain on music: The science of a human obsession*. New York: Plume.
- Lindenberger, U., Li, S., Gruber, W., & Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC neuroscience*, 10(22). doi: 10.1186/1471-2202-10-22
- London, J. (2012). *Hearing in time: Psychological aspects of musical meter* (2nd ed.). New York, NY: Oxford University Press.
- London, J., Himberg, T., & Cross, I. (2009). The effect of structural and performance factors in the perception of anacruses. *Music Perception*, 27(2), 103–120.
- Lucas, G., Clayton, M., & Leante, L. (2011). Inter-group entrainment in afro-brazilian congado ritual. *Empirical Musicology Review*, 6(2), 75–102.
- Luck, G., & Nte, S. (2008). An investigation of conductors’ temporal gestures and conductor—musician synchronization, and a first experiment. *Psychology of music*, 36(1), 81–99.
- Luck, G., Riikkilä, K., Lartillot, O., Erkkilä, J., Toiviainen, P., Mäkelä, A., ... Värri, J. (2006). Exploring relationships between level of mental retardation and features of music therapy improvisations: A computational approach. *Nordic Journal of Music Therapy*, 15(1), 30–48.
- Luck, G., & Toiviainen, P. (2006). Ensemble musicians’ synchronization with conductors’ gestures: An automated feature-extraction analysis. *Music Perception*, 24(2), 189–200.
- Luck, G., Toiviainen, P., Erkkilä, J., Lartillot, O., Riikkilä, K., Mäkelä, A., ... Värri, J. (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music*, 36(1), 25–45.
- MacDonald, R., Kreutz, G., & Mitchell, L. (2012). *Music, health, and wellbeing*. Oxford, UK: OUP Oxford.

- MacDougall, H. G., & Moore, S. T. (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of applied physiology*, 99(3), 1164–1173. doi: 10.1152/japplphysiol.00138.2005
- Macrae, C., & Johnston, L. (1998). Help, i need somebody: Automatic action and inaction. *Social Cognition*, 16(4), 400–417.
- Madison, G. (2004). Detection of linear temporal drift in sound sequences: empirical data and modelling principles. *Acta Psychologica*, 117(1), 95–118.
- Maduell, M., & Wing, A. M. (2007). The dynamics of ensemble: the case for flamenco. *Psychology of Music*, 35(4), 591–627. doi: 10.1177/0305735607076446
- Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality - exploring the basis of human companionship* (pp. 1–11). Oxford University Press.
- Marsh, K., Johnston, L., Richardson, M., & Schmidt, R. (2009). Hop off the mirror neuron bandwagon and join ours, it's less crowded! *European Journal of Social Psychology*, 39(7), 1234–1235. doi: 10.1002/ejsp.689
- Marsh, K. L., Johnston, L., Richardson, M. J., & Schmidt, R. C. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology*, 39(7), 1217–1225. (There are also commentaries and a reply linked to this main article.) doi: 10.1002/ejsp.666
- Marwan, N. (2012). *Cross recurrence plot toolbox 5.16*. <http://tocsy.pik-potsdam.de/CRPtoolbox/>.
- Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5–6), 237 - 329. doi: DOI: 10.1016/j.physrep.2006.11.001
- Marwan, N., Thiel, M., & Nowaczyk, N. R. (2002). Cross recurrence plot based synchronization of time series. *Nonlinear Processes in Geophysics*, 9(3/4), 325–331. doi: 10.5194/npg-9-325-2002
- Mates, J. (1994). A model of synchronization of motor acts to a stimulus sequence i. *Biological cybernetics*, 70, 463–473.
- Mates, J., & Aschersleben, G. (2000, 2000/3). Sensorimotor synchronization: the impact of temporally displaced auditory feedback. *Acta Psycho-*

- logica*, 104(1), 29-44. (TY - JOUR)
- Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, 6(4), 332-340.
- Mates, J., Radil, T., & Pöppel, E. (1992, December). Cooperative tapping: time control under different feedback conditions. *Perception & Psychophysics*, 52(6), 691-704.
- Mattar, A., & Gribble, P. (2005). Motor learning by observing. *Neuron*, 46(1), 153-160. doi: 10.1016/j.neuron.2005.02.009
- McArthur, L., & Baron, R. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90(3), 215-238. doi: 10.1037/0033-295X.90.3.215
- McAuley, J. D. (1995). *On the perception of time as phase: toward an adaptive-oscillator model of rhythm*. Phd thesis, Indiana University, Bloomington, IN.
- McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, 135(3), 348-367. doi: 10.1037/0096-3445.135.3.348
- McNeill, W. H. (1995). *Keeping together in time*. Harvard University Press.
- Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7(274). doi: 10.3389/fnins.2013.00274
- Merleau-Ponty, M. (2002 (1962)). *Phenomenology of perception*. London, UK: Routledge.
- Moiseff, A., & Copeland, J. (2000). A new type of synchronized flashing in a north american firefly. *Journal of insect behavior*, 13(4), 597-612. doi: 10.1023/A:1007823904866
- Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to 'e'motion. *Soc Cogn Affect Neurosci*, 1(3), 235-241. doi: 10.1093/scan/nsl029
- Moore, G., & Chen, J. (2010). Timings and interactions of skilled musicians. *Biological Cybernetics*, 103, 401-414. doi: 10.1007/s00422-010-0407-5
- Murningham, J., & Conlon, D. (1991). The dynamics of intense work groups: a study of british string quartets. *Administrative Science Quarterly*,

- 36, 165–186.
- Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., & Ponton, C. (2011). The mismatch negativity (mmn)—a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clinical Neurophysiology*, 123, 424–258.
- Nessler, J. A., & Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Human Movement Science*, 28, 772–785. doi: 10.1016/j.humov.2009.04.007
- Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 463–472). MIT Press.
- Nowak, A., & Vallacher, R. R. (1998). Toward computational social psychology: Cellular automata and neural network models of interpersonal dynamics. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 277–311). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Nowicki, L., Prinz, W., Grosjean, M., Repp, B. H., & Keller, P. E. (2013). Mutual adaptive timing in interpersonal action coordination. *Psychomusicology: Music, Mind, and Brain*, 23(1), 6–20. doi: 10.1037/a0032039
- Noy, L., Dekel, E., & Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences*, 108(52), 20947–20952. doi: 10.1073/pnas.1108155108
- Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, I. P., Hari, R., & Sams, M. (2012, May). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1206095109
- Okubo, A. (1986). Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in biophysics*, 22, 1–94.
- Oullier, O., de Guzman, G., Jantzen, K., Lagarde, J., & Kelso, J. (2008). Social coordination dynamics: Measuring human bonding. *Social neuroscience*, 3(2), 178–192. doi: 10.1080/17470910701563392
- Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, 26(5),

489–504. doi: 10.1525/MP.2009.26.5.489

- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of experimental psychology: human perception and performance*, 15(2), 331–346.
- Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48.
- Palmer, C., & Meyer, R. (2000). Conceptual and motor learning in music performance. *Psychological Science*, 11(1), 63–68.
- Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, 12(1).
- Papiotis, P., Marchini, M., & Maestre, E. (2013). Multidimensional analysis of interdependence in a string quartet. In A. Williamon & W. Goebel (Eds.), *Proceedings of the international symposium on performance science 2013, vienna*. European Association of Conservatoires.
- Patel, A. D. (2010). *Music, language, and the brain*. New York: Oxford University Press.
- Patel, A. D., & Iversen, J. (2006). A non-human animal can drum a steady beat on a musical instrument. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th international conference on music perception and cognition* (p. 477).
- Patel, A. D., Iversen, J., Bregman, M., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current biology*, 19(10), 827–830. doi: 10.1016/j.cub.2009.03.038
- Pavlicevic, M. (1997). *Music therapy in context - music, meaning and relationship*. London, UK: Jessica Kingsley Publishers, Ltd.
- Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental brain research*, 211, 1–11. doi: 10.1007/s00221-011-2616-0
- Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1), 176–180.
- Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., & Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron*, 33(2), 185–191.
- Peretz, I., Champod, A., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, 999(1), 58–75.

- Perlovsky, L. (2010). Musical emotions: Functions, origins, evolution. *Physics of Life Reviews*, 7(1), 2–27.
- Pfordresher, P., & Palmer, C. (2006). Effects of hearing the past, present, or future during music performance. *Attention, Perception, & Psychophysics*, 68(3), 362–376.
- Phillips-Silver, J., Aktipis, C., & Bryant, G. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music perception*, 28(1), 3–14.
- Phillips-Silver, J., Toivainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, 49(5), 961–969.
- Phillips-Silver, J., & Trainor, L. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, 308(5727), 1430. doi: 10.1126/science.1110922
- Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, 105(3), 533 – 546. doi: DOI: 10.1016/j.cognition.2006.11.006
- Piaget, J. (1952). *The origin of intelligence in children*. WW Norton Co.
- Pike, K. L. (1967). Etic and emic standpoints for the description of behavior. *Language and thought: an enduring problem in psychology*, 37, 32.
- Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization - a universal concept in non-linear sciences* (No. 12). Cambridge: Cambridge University Press.
- Povel, D. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance; Journal of Experimental Psychology: Human Perception and Performance*, 7(1), 3–18.
- Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological cybernetics*, 76(5), 339–347.
- Prinz, W. (1997). Perception and action planning. *European journal of cognitive psychology*, 9(2), 129–154.
- Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2.5-and 4-year-old children. *International Journal of Behavioral Development*, 27(3), 220–231.
- Rabinowitch, T.-C., Cross, I., & Burnard, P. (2012). Long-term musical group interaction has a positive influence on empathy in children. *Psy-*

chology of Music.

- Rasch, R. (1979). Synchronization in performed ensemble music. *Acustica*, 43, 121–131.
- Rasch, R. (1988). Timing and synchronization in ensemble performance. In J. Sloboda (Ed.), *The generative processes in music* (p. 70-90). Oxford: University Press.
- Renvall, V., & Malinen, S. (2012). Setup and apparatus for two-person fMRI (poster presentation). In *Human Brain Mapping 2012*. Beijing, China.
- Repp, B. H. (2005a). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 1. qualitative observations. *Music Perception*, 22(3), 479-496. (Journal Article)
- Repp, B. H. (2005b). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6), 969-992.
- Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science*, 27(3), 423-456.
- Repp, B. H., London, J., & Keller, P. E. (2012). Distortions in Reproduction of Two-Interval Rhythms: When the “Attractor Ratio” is not exactly 1 : 2. *Music Perception*, 30(2), 205–223.
- Repp, B. H., & Penel, A. (2002, 2002/10). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences1. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1085-1099. (TY - JOUR)
- Repp, B. H., & Penel, A. (2004, 2004/8). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, 68(4), 252-270. (TY - JOUR)
- Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic bulletin & review*, 20, 403–452. doi: 10.3758/s13423-012-0371-2
- Richardson, M., Marsh, K., Isenhower, R., Goodman, J., & Schmidt, R. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891. doi: 10.1016/j.humov.2007.07.002
- Richardson, M., Marsh, K., & Schmidt, R. (2010). Challenging the egocentric view of coordinated perceiving, acting and knowing. In L. Barrett, B. Mesquita, & E. Smith (Eds.), *The mind in context* (pp. 307–333).

New York, NY: The Guilford Press.

- Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 62-79.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2), 131-141.
- Robertson, D. G. E., Caldwell, G. E., Hamill, J., Kamen, G., & Whittlesley, S. N. (2004). *Research methods in biomechanics*. Champaign, IL: Human Kinetics.
- Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. a critical review. *International Journal of Psychophysiology*(0), -. doi: 10.1016/j.ijpsycho.2011.12.010
- Ruud, E. (2010). *Music therapy: A perspective from the humanities*. Gilsum, NH: Barcelona Publishers.
- Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, 35(1), 88-109.
- Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Hupertz, H.-J., ... Schulze-Bonhage, A. (2012). Co-localizing linguistic and musical syntax with intracranial eeg. *NeuroImage*(0), -. doi: 10.1016/j.neuroimage.2012.09.035
- Sänger, J., Lindenberger, U., & Muller, V. (2011). Interactive brains, social minds. *Communicative & Integrative Biology*, 4(6), 655-663. doi: 10.4161/cib.4.6.17934
- Sänger, J., Muller, V., & Lindenberger, U. (2012). Intra- and inter-brain synchronization and network properties when playing guitar in duets. *Frontiers in Human Neuroscience*, 6(312). doi: 10.3389/fnhum.2012.00312
- Sawyer, K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music*, 34(2), 148-165. (Collection of papers from Musical Collaboration) doi: 10.1177/0305735606061850
- Schachner, A., Brady, T., Pepperberg, I., & Hauser, M. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19(10), 831-836.
- Schellenberg, E. G. (2004). Music lessons enhance iq. *Psychological Science*,

15(8), 511-514.

- Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, 1060(1), 219-230.
- Schmidt, R., Bienvenu, M., Fitzpatrick, P., & Amazeen, P. (1998). A comparison of intra-and interpersonal interlimb coordination: Coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 884.
- Schmidt, R., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (Vol. 17, p. 281-308). Berlin: Springer Verlag.
- Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, 9, 189-206.
- Schmidt, R. C., & Turvey, M. T. (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biological cybernetics*, 70(4), 369-376.
- Schulze, H.-H., & Vorberg, D. (2002, 2002/2). Linear phase correction models for synchronization: Parameter identification and estimation of parameters1. *Brain and Cognition*, 48(1), 80-97. (TY - JOUR)
- Sebanz, N., Bekkering, G., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Science*, 10(2), 70-76.
- Sebanz, N., & Knoblich, G. (2009). Jumping on the ecological bandwagon? mind the gap! *European Journal of Social Psychology*, 39(7), 1230-1233. doi: 10.1002/ejsp.688
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, 88(3), B11-B21. doi: 10.1016/S0010-0277(03)00043-X
- Semin, G. R., & Cacioppo, J. T. (2008). Grounding social cognition: Synchronization, coordination, and co-regulation. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding - social, cognitive, affective, and neuroscientific approaches* (p. 119-147). Cambridge University Press.
- Senoga-Zake, G. W. (2000). *Folk music of kenya*. Nairobi, Kenya: Uzima Press.
- Seth, A. (2007). Granger causality. *Scholarpedia*, 2(7), 1667. doi:

- 10.4249/scholarpedia.1667
- Seth, A. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of neuroscience methods*, 186(2), 262–273. doi: 10.1016/j.jneumeth.2009.11.020
- Shaffer, H. (1984). Timing in musical performance. *Annals of the New York Academy of Sciences*, 423(1), 420–428. doi: 10.1111/j.1749-6632.1984.tb23450.x
- Shapiro, L. (2007). The embodied cognition research programme. *Philosophy compass*, 2(2), 338–346. doi: 10.1111/j.1747-9991.2007.00064.x
- Shapiro, L. (2011). *Embodied cognition*. Abingdon, UK: Routledge.
- Shockley, K. (2005). Cross recurrence quantification of interpersonal postural activity. In M. A. Riley & G. C. van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 142–177). Retrieved 28 September, 2012 from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>.
- Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 201–208. doi: 10.1037/0096-1523.33.1.201
- Shockley, K., Butwill, M., Zbilut, J. P., & Webber Jr, C. L. (2002). Cross recurrence quantification of coupled oscillators. *Physics Letters A*, 305(1), 59–69. doi: 10.1016/S0375-9601(02)01411-1
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1, 305–319. doi: 10.1111/j.1756-8765.2009.01021.x
- Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology-Human Perception and Performance*, 29(2), 326–332.
- Shumway, R. H. (1988). *Applied statistical time series analysis* (R. A. Johnson & D. W. Wichern, Eds.). Englewood Cliffs, New Jersey: Prentice Hall.
- Small, C. (1998). *Musicking: The meanings of performing and listening*. Middletown, CT: Wesleyan University Press.
- Spiro, N., & Himberg, T. (2012). Musicians and non-musicians adapting to tempo differences in cooperative tapping tasks. In E. Cambouropoulos,

- C. Tsourgas, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th international conference on music perception and cognition and the 8th triennial conference of the european society for the cognitive sciences of music*. Thessaloniki, Greece.
- Standley, J. M. (2002). A meta-analysis of the efficacy of music therapy for premature infants. *Journal of Pediatric Nursing*, 17(2), 107 - 113. doi: 10.1053/jpnd.2002.124128
- Stephan, K., Thaut, M., Wunderlich, G., Schicks, W., Tian, B., Tellmann, L., ... Hömberg, V. (2002). Conscious and subconscious sensorimotor synchronization—prefrontal cortex and the influence of awareness. *Neuroimage*, 15(2), 345–352.
- Stevens, L. (1886). On the time-sense. *Mind*, 11(43), 393–404.
- Stewart, L. (2006). Congenital amusia. *Current Biology*, 16(21), 904–906.
- Stobart, H., & Cross, I. (2000). The andean anacrusis? rhythmic structure and perception in easter songs of northern potosi, bolivia. *British Journal of Ethnomusicology*, 9(2), 63–92.
- Strogatz, S. (2003). *Sync: The emerging science of spontaneous order*. New York, NY: Hyperion.
- Summers, J. J., Todd, J. A., & Kim, Y. H. (1993). The influence of perceptual and motor factors on bimanual coordination in a polyrhythmic tapping task. *Psychological Research*, 55(2), 107–115.
- Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence, warwick 1980* (Vol. 898, p. 366–381). Springer Berlin / Heidelberg.
- Temperley, D. (2000). Meter and grouping in african music: A view from music theory. *Ethnomusicology*, 44(1), 65–96.
- Thaut, M. H. (2008). *Rhythm, music, and the brain: Scientific foundations and clinical applications*. New York: Routledge.
- Thaut, M. H., Gardiner, J. C., Holmberg, D., Horwitz, J., Kent, L., Andrews, G., ... McIntosh, G. R. (2009). Neurologic music therapy improves executive function and emotional adjustment in traumatic brain injury rehabilitation. *Annals of the New York Academy of Sciences*, 1169(1), 406–416.
- Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge MA: MIT Press.
- Thiel, M., Romano, M., Kurths, J., Meucci, R., Allaria, E., & Arecchi, F.

- (2002). Influence of observational noise on the recurrence quantification analysis. *Physica D: Nonlinear Phenomena*, 171(3), 138–152.
- Thompson, M. R. (2012). *The application of motion capture to embodied music cognition research*. Phd thesis, Department of Music, University of Jyväskylä, Jyväskylä, Finland.
- Todd, N. P. M., Lee, C. S., & O’Boyle, D. J. (2002, 2002/2). A sensorimotor theory of temporal tracking and beat induction. *Psychological Research*, 66(1), 26–39. (TY - JOUR)
- Todd, N. P. M., O’Boyle, D. J., & Lee, C. S. (1999). A sensory-motor theory of rhythm, time perception and beat induction. *Journal of New Music Research*, 28(1), 005 - 028.
- Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. A. S. (2007, May). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences*, 104(19), 8190–8195.
- Toiviainen, P. (1998). An interactive midi accompanist. *Computer Music Journal*, 22(4), 63–75.
- Toiviainen, P., & Burger, B. (2010). *Mocap toolbox manual* (Vol. Available at <http://www.jyu.fi/music/coe/materials/mocaptoolbox/MCTmanual>). Jyväskylä, Finland: University of Jyväskylä.
- Toiviainen, P., & Eerola, T. (2003). Where is the beat?: Comparison of finnish and south african listeners. In R. Kopiez, A. C. Lehmann, I. Wolther, & C. Wolf (Eds.), *5th triennial escom conference* (p. 501–504). Hanover, Germany: Hanover University of Music and Drama.
- Toiviainen, P., Luck, G., & Thompson, M. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception*, 28, 59–70. doi: 10.1525/mp.2010.28.1.59
- Tolonen, T., & Karjalainen, M. (2000). A computationally efficient multipitch analysis model. *IEEE Transactions on Speech and Audio Processing*, 8(6), 708–716.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121–125. doi: 10.1111/j.1467-7687.2007.00573.x
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–735.
- Toussaint, G. T. (2013). *The geometry of musical rhythm: What makes a “good” rhythm good?* Boca Raton, FL: CRC Press.

- Trainor, L. J., Gao, X., Lei, J.-j., Lehtovaara, K., & Harris, L. R. (2009). The primal role of the vestibular system in determining musical rhythm. *Cortex*, 45(1), 35–43. doi: 10.1016/j.cortex.2007.10.014
- Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development*, 21(1), 77–88.
- Trehub, S. E. (2003, 2003/7). The developmental origins of musicality. *Nature Neuroscience*, 6(7), 669–673. (TY - JOUR)
- Trevarthen, C. (2006). Reflections on the nature and emotional needs of infant intersubjectivity. In C. S. Carter et al. (Eds.), *Attachment and bonding: A new synthesis* (Vol. 92, pp. 55–84). MIT Press.
- Turino, T. (2008). *Music as social life: The politics of participation*. Chicago, IL: University of Chicago Press.
- Vallacher, R., & Jackson, D. (2009). Thinking inside the box—dynamical constraints on mind and action: Comment on marsh et al.’s toward a radically embodied, embedded social psychology, this issue. *European Journal of Social Psychology*, 39(7), 1226–1229. doi: 10.1002/ejsp.687
- Vallacher, R., & Nowak, A. (2002). The dynamical perspective in personality and social psychology. *Personality and Social Psychology Review*, 6(4), 264–273. doi: 10.1207/S15327957PSPR060401
- van Baaren, R., Holland, R., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, 15(1), 71.
- Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.
- Vorberg, D. (2005). Synchronization in duet performance. In *10th rhythm perception and production workshop*. Bilzen, Belgium.
- Vorberg, D., & Schulze, H.-H. (2002). Linear phase-correction in synchronization: Predictions, parameter estimation, and simulations1. *Journal of Mathematical Psychology*, 46(1), 56–87.
- Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S. Keele (Eds.), *Handbook of perception and action, vol. 2, motor skills* (p. 181–262). London: Academic Press.
- Vuoskoski, J., Thompson, W., McIlwain, D., & Eerola, T. (2012). Who enjoys listening to sad music and why? *Music Perception*, 29(3), 311–317.
- Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard,

- L. (2006). It don't mean a thing....: Keeping the rhythm during polyrhythmic tension, activates language areas (ba47). *Neuroimage*, 31(2), 832–841. doi: 10.1016/j.neuroimage.2005.12.037
- Wallin, N. L., Merker, B. H., & Brown, S. (Eds.). (2000). *The origins of music*. Cambridge, MA: MIT Press.
- Wanderley, M., Vines, B., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34(1), 97–113.
- Webber, C. L., Marwan, N., Facchini, A., & Giuliani, A. (2009). Simpler methods do it better: Success of recurrence quantification analysis as a general purpose data analysis tool. *Physics Letters A*, 373(41), 3753 – 3756. doi: DOI: 10.1016/j.physleta.2009.08.052
- Wigram, T., & Gold, C. (2006). Music therapy in the assessment and treatment of autistic spectrum disorder: clinical application and research evidence. *Child: Care, Health and Development*, 32(5), 535–542. doi: 10.1111/j.1365-2214.2006.00615.x
- Williamon, A., & Davidson, J. (2002). Exploring co-performer communication. *Musicae Scientiae*, 1(1), 53–72.
- Williams, K. D. (2010). Dyads can be groups (and often are). *Small Group Research*, 41(2), 268–274. doi: 10.1177/1046496409358619
- Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in psychology*, 4(58), 1–13. doi: 10.3389/fpsyg.2013.00058
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. doi: 10.1.1.124.8295
- Wilson, M., & Wilson, T. (2005). An oscillator model of the timing of turn-taking. *Psychonomic bulletin & review*, 12(6), 957–968.
- Wiltermuth, S. S., & Heath, C. (2009, Jan). Synchrony and cooperation. *Psychological Science*, 20(1), 1–5. doi: 10.1111/j.1467-9280.2008.02253.x
- Wing, A. M., & Kristofferson, A. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, 14, 5–12.
- Wing, A. M., & Woodburn, C. (1995). The coordination and consistency of rowers in a racing eight. *Journal of Sports Sciences*, 13, 187–197.
- Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, 106(7), 2468–2471. doi:

10.1073/pnas.0809035106

- Wundt, W. M. (1904). *Principles of physiological psychology*. Sonnenschein.
- Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports*, 2(959). doi: 10.1038/srep00959
- Zarco, W., Merchant, H., Prado, L., & Mendez, J. (2009). Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology*, 102(6), 3191–3202.
- Zbilut, J., Giuliani, A., & Webber, C. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, 246(1), 122–128.
- Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 107(13), 5768–5773.

Appendix A

Pilot experiments and student projects

A.1 Introduction

In this appendix, I will briefly describe some of the small pilot studies and student projects that I have supervised on cooperative tapping. Instead of presenting their results in full, I will just explain their hypotheses and methods and briefly discuss the main results, if there were any. These results are tentative and due to both the small number of participants and many methodological issues, they should be taken with many grains of salt. These small studies (typically $6 < n < 12$) are under-powered, so even with p-values below the commonly used threshold of 0.05, the likelihood of the results being *correct* would be rather small. This problem of underpowered, small studies has recently been discussed in detail in the fields of biomedical research and neuroscience (Button et al., 2013; Ioannidis, 2005).

A.2 Auditory and Visual Modalities

One familiar strategy in studies of interpersonal coordination is to compare coordination using the visual or auditory modalities (Mates et al., 1992). From previous studies it is known that for synchronisation, the auditory rhythms are more engaging and that sensorimotor synchronisation is more accurate with an auditory target than a visual one (Repp & Penel, 2002, 2004). The roles of the auditory and visual communication channels in co-

ordinating interpersonal entrainment had not been investigated, however. Intuitively, the visual modality would seem to be important for coordination mostly at the starts and finishes of a piece, and perhaps at pivotal points and junctions of musical structure, where note onsets can not be predicted with adequate accuracy based on prior context. In the much simpler tapping task, and in the absence of such junctions, the auditory modality would probably dominate, as it contains all the task-critical information. However, how much better would the auditory modality be compared to the visual one, and how would the naturalistic auditory-visual condition fare?

The comparison of auditory, visual and auditory plus visual communication channels featured in a number of small pilots and student projects. The tapping tasks were synchronisation-continuation tasks, along with a combination of other factors (auditory delays, in-phase and anti-phase synchronisation, different effectors, see below). In the auditory only condition (A), the participants could hear each other's taps but not see each other. In the visual only condition (V), they could see but not hear each other. In the auditory plus visual condition (AV), they could both see and hear each other. Participants heard feedback for their own taps in all conditions.

In all but one of the pilot studies, pairs achieved higher levels of synchrony in the auditory only condition than in the visual only condition. This was an expected result, in line with previous research on modalities (Repp & Penel, 2002, 2004). The results regarding the auditory and visual condition have been less clear. In some pilots, participants are as synchronised in the auditory plus visual condition as they are in the auditory only condition, in others, there is a clear difference to the advantage of the auditory only condition. This inconsistency is probably explained by differences in the other experimental factors.

A typical result is the one in Rebecca Saxby's study on assertiveness (see full description of the experiment below, and results in figure A.1). Here, participants synchronised with the metronome best (lowest average asynchrony between their and the metronome's onsets) when they could only see the other participant, and their partner's tap sounds were not disturbing them. However, looking at coordination with the partner (average asynchronies between the partners' onsets), auditory and auditory + visual have the lowest asynchronies.

Interestingly, many participants preferred to not look at their partner

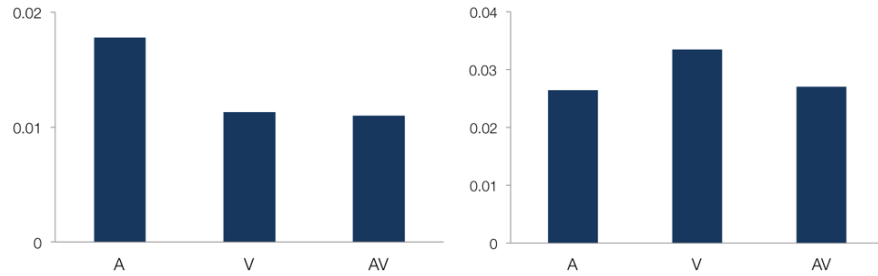


Figure A.1: Synchronisation (left panel) and coordination errors (right panel) in a tapping trial (measured in seconds), in auditory only (A), visual only (V), and auditory-visual (AV) conditions.

even when that was possible. Thus the AV condition was often reduced to just A. Some participants focused their eyes to the touch pad, some were even shading their eyes with their free hand, to block their view to the other person. This is not terribly surprising, as all the task-relevant information is in the auditory channel. Perhaps these participants wanted to focus as much of their attention as possible to this information and block out all potentially disturbing sensory inputs.

On the other hand, many participants utilised the visual channel to communicate with their partner. For example, a few participants started to exaggerate their hand movements in order to stop their partner drifting out of sync or out of what they considered the correct tempo. Few even started to conduct their partner by signing the beat subdivisions. These are clear examples of one participant taking the role of a leader in the interaction.

In most musical interactions most of the coordination is based on listening to the other players. The visual modality is important mainly for coordinating the starts and stops and the various junctions in the musical texture. In ensembles, the role of signalling the timing is usually delegated to the leader of the ensemble. Figuring out how to coordinate and communicate at various points of a musical piece is a very important part of learning to perform it as an ensemble.

A.3 Delayed auditory feedback

The effects of auditory delay on synchronisation accuracy have been studied in solo settings (Mates & Aschersleben, 2000; Pfordresher & Palmer, 2006). In a student project by Sam Hudson, the effect of delayed auditory feedback (DAF) on synchronisation and coordination accuracy was investigated. The student conducting the study was an organ scholar, and therefore used to playing in conditions where there would be a long delay between pressing the organ’s keys and hearing the output. Coordinating with the chapel choir under such conditions is challenging, but a skill that organists are able to learn.

In the experiment, 12 music students performed dyadic tapping tasks, having either auditory or visual connection or both (A, V, and AV conditions, respectively). They would tap isochronous crotchets, or a simple rhythmic pattern in quavers and crotchets. These tasks were conducted with and without a 50 millisecond auditory delay added to the outputs of the participant tapping the quaver pattern. The tapping trials were paced by an auditory metronome.

Unsurprisingly, adding a 50 ms delay increases the average synchronisation error between the delayed tapper and the metronome. The auditory only condition (A) was influenced the most, and the AV condition the least. Looking at entrainment between participants, the delay increased the average asynchrony slightly in the V and AV conditions, and very clearly in the A condition. There were no significant differences between the conditions when there were no delays. The A condition was the best but not statistically significantly better than the two others. The lack of difference between the A and V conditions can be explained by the auditory metronome that was heard by the participants also in the V condition, where they could not hear each other’s tapping.

Interestingly, the asynchronies here were again smaller between the participants (coordination / entrainment) than between the participants and the metronome. Similar results were observed in many of the pilot studies.

A.4 Metrical Schemata

In the cognitive theories of interpersonal interaction, a key aspect in joint actions is the alignment of mental representations (Garrod & Pickering, 2009; Tomasello & Carpenter, 2007). If we assume mental representations exist (see discussion in chapter 2.4.6), in musical interactions, metric schemata are a strong candidate for such representations. How having the same or different schemata influences entrainment would help shed light to this question.

Participants often group beats into groups of two, four, or more rarely three, even in isochronous tapping tasks. According to post-experiment interviews, this helps them maintain a steady tempo. Instead of a “flat” isochronous beat they evoke a metrical structure for their tapping, often accentuating the metrically strong beats or tapping their feet to every other or every fourth beat. In a student project by Dan Soper, the effects of matching or mixed metrical schemata were examined. The idea was to instruct each of the two participants to adopt either a $\frac{3}{4}$ or a $\frac{4}{4}$ schema, without telling the other participant. We would then measure their entrainment accuracy in all four different combinations.

Twelve music students were recruited, and they performed these dyadic finger tapping tasks in the already familiar A, V and AV conditions. There were both synchronisation and synchronisation-continuation trials. Each trial was started by five metronome tones, so that the metronome would not evoke either of the metrical schemata the participants were supposed to adopt.

The tempo (crotchet length) was held constant, so that the event rate was the same for both participants. Thus, the participants were synchronised at the beat level but not at the bar level (in other words, their strong and weak beats were not synchronised).

The results of this experiment were somewhat confusing and at parts difficult to interpret, and we can not rule out human error in the process. In such a small study, and using average (unsigned) asynchronies as the outcome measure, the results can also be skewed quite a lot by an outlier performance by an individual pair in an individual trial.

This study was the one exception where synchronisation in the synchronisation tasks with the metronome and with the partner were best in the visual only condition. The result of the synchronisation-continuation tasks

was comparable with the other pilot studies in that the A and AV conditions turned out to be best. The auditory metronome in the synchronisation tasks can at least partially explain this result. The V condition is the only one where the partner’s tapping does not interfere with the task of synchronising with the metronome.

Furthermore, with regard to the metrical schemata conditions, smallest asynchrony with the metronome was obtained when both participants were tapping in $\frac{3}{4}$ time. This is surprising, as the hypothesis was that the more familiar $\frac{4}{4}$ would top the list. It is possible that this is explained by participants paying more attention in this less familiar condition. Alternatively, it might be an effect of the faster high-level period in the $\frac{3}{4}$ condition. As predicted, the mixed conditions yielded the largest synchronisation errors. This prediction was based on the hypothesis that the participants would have most trouble in synchronising when there were de-synchronised at the bar level.

In the continuation tasks, results flip again, with the $\frac{3}{4}$ condition yielding higher asynchronies than the other conditions. Thus, the role of metrical schemata remains unclear after this study.

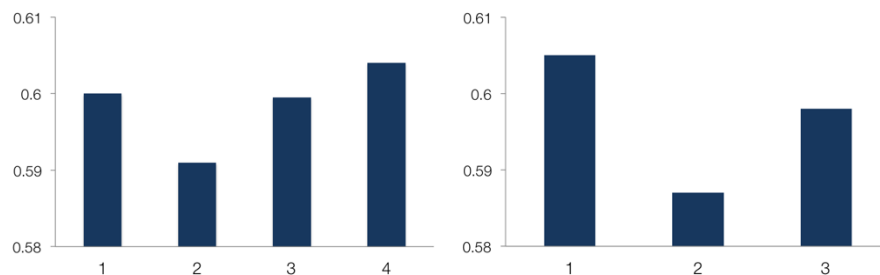


Figure A.2: Average inter-tap intervals of the beats in the bar for one pair. $\frac{4}{4}$ in the left panel, $\frac{3}{4}$ in the right one.

The implementation of the metrical schemata was interesting, however. The patterns of strong and weak beats were implemented in both the timing (the only information available for the participants) as well as in the force of the taps. The timings of the beats followed the patterns in figure A.2. Strong beats and the last beat of the bar are longer than the weak beats. Even though the system played back all taps at the same (or similar) loudness,

the different forces were still recorded by the software. One pair’s data was selected for this analysis. In terms of tapping force, this study suggests that even tappers who are fairly inaccurate in their timing are good at assigning the beats of a bar different weights by playing strong beats with a greater force than the weak beats¹. The force profiles of these tappers, who were instructed to tap either in three or in four, had clear autocorrelations at the predicted lags. Furthermore, these profiles correlated very significantly with the metrical accents as described by Lerdahl and Jackendoff (1983). These correlations are in the table A.1.

Table A.1: Correlations of relative velocities and metrical hierarchies

	Tapping in $\frac{3}{4}$	Tapping in $\frac{4}{4}$
GTTM $\frac{3}{4}$.969(*)	−.007
GTTM $\frac{4}{4}$.873(*)	.010

The role of metric schemata should be investigated in a properly executed and preferably a little larger study. The basic design could be copied from this pilot, although I would no longer divide the synchronisation and continuation tasks as separate trials, but would rather conduct combined synchronisation-continuation trials that would be longer than the ones used in this study. Also, even though tapping force was not supposed to influence the loudness of the output sounds, this was implemented by selecting the flattest output curve from the drum. In a proper experiment, this would need to be implemented in the synthesiser to make sure the outputs are definitely identical every time.

A.5 Effects of personality: assertiveness

How does personality effect musical interaction? For chamber musicians, playing together is easier with some partners than others. Would this be reflected at the level of entrainment, or are the personality effects limited

¹MIDI quantifies the force of the drum beat, piano key press etc. as “velocity”. This refers to the downward velocity of a piano key. It is a relative measure, ranging from 0 to 127, and the absolute values are not directly related to actual, physical velocity of the finger or hand.

to “higher level” interactions such as artistic decision-making (Himberg & Spiro, 2012)? In a student project with Rebecca Saxby, participants filled in assertiveness questionnaires (SIB, (Arrindell & van der Ende, 1985)) and then performed dyadic finger tapping tasks.

The results regarding synchronisation and coordination accuracy in the various sensory modalities are plotted in the previous section in figure A.1. The only surprise here is that the asynchronies to the metronome were smaller than those to the other participant. The main result of the experiment, correlations between the synchronisation and coordination measures and the assertiveness scores turned out to be negative.

Anecdotally, the pair that had the closest coordination and most equal performance was also the one that had the largest difference in assertiveness scores. Their differences in non-musical, social behaviour were noticeable even in the short time they were in the lab, but their musical interaction was extremely well-balanced and skillful. This was an interesting case that lead us to discuss how musical interaction differs from verbal interaction. Musical performance provides an alternative means of expression to people who might have difficulties in verbal social interactions. Music does not operate on propositional meaning but rather gives a lot of flexibility to meanings; it has “floating intentionality”(I. Cross, 2008). Also, it is typical in Western classical music that music lessons are given in one-to-one situations, and these skills are then developed by practicing in privacy. This setting allows or even favours the more introverted personalities (Kemp, 1981). The efficacy of music therapy for those in the autistic spectrum probably is also linked to the safety and flexibility of musical interactions.

A.6 Synchronisation and syncopation - effector comparison

In a classic study, Scott Kelso and collaborators observed that at slow tempi, bimanual coordination (finger-wiggling) was stable both in-phase and anti-phase. However, when the rate of finger movements was increased to above 6 Hz, the anti-phase relationship no longer could be maintained, and the movements shifted into an in-phase arrangement (Haken et al., 1985; Kelso, 1984). Coordination between two people was found to follow similar dynamics (R. Schmidt et al., 1998).

Using simpler measurement methods, we conducted a dyadic tapping study that investigated the effect of tempo to entrainment across different tempi. This was a project by two students, Kathryn Loxley and Alice Williamson. In this study, 10 music students took part. Fingertapping was compared to tapping with a drum stick. Synchronisation and syncopation tasks were conducted in pairs with a metronome that started relatively slow (750 ms inter-onset interval), and stepped up to a faster tempo every 12 beats, until after 9 steps it reached the IOI of 233 ms.

Participants were more accurate with their fingers than with drum sticks. When the metronome stepped up to a new, faster tempo, participants overcompensated at first, speeding past the metronome first before settling to the new tempo. Just as in the studies by Kelso and Schmidt et al., there was very little difference in performance in slower tempi, but tapping on the beat (in-phase synchrony) was more accurate in faster tempi than syncopating (anti-phase synchrony).

Again, in this study there were too many factors, which decreases the quality of the data and makes it more difficult to say anything conclusive about the main factor. Interestingly, the pair that performed best in anti-phase synchrony had a “sneaky” strategy: both participants imagined they were the ones tapping on beat. This allowed them to get through even the fastest tempo level without the rhythm breaking down.

A.7 Autism and entrainment

In a student project by Camilla Farrant, data were collected from participants with autistic spectrum disorders. The initial plan was to do a small tapping study, where we would compare tapping with music, with a metronome, and with another person.

We got usable data from three participants, one boy and two girls. None of them completed our planned trials, however. The two girls in the group were the most consistent tappers, but both had their own tempo that they maintained regardless of what the metronome, the music, or the experimenter were doing. They both demonstrated excellent skills in maintaining a steady tempo and in resisting entrainment. On the fly, we decided that Camilla should try to tap in synchrony in whatever tempo they chose, and so we got some data that was presented in chapter 6.7 on page 152 as an example of

using windowed cross-correlation to visualise mutual adaptation. The plots in figure 6.4 indicate that Camilla was the one adapting while our participant kept going in her own, steady tempo.

This was my first introduction to the world of autism. For social interaction and communication research, autism is of course highly relevant, as it is often characterised by impairments in social behaviours. One of the main lessons I learned from this pilot and other studies is that if we want to study (musical) interaction within a wide range of participants in the autism spectrum, and not just among those with high-functioning Asperger’s syndrome, any research effort must be embedded in regular music therapy sessions, and the researcher must not expect the interactions to have much musical structure or repeatability. As a research topic it is a treasure trove, however, and working with music therapists is important both for being able to design better studies and for having a direct feedback channel to get the research results into practice, hopefully contributing to improvement of therapeutic methods.

A.8 Summary

In general, these pilot studies were great opportunities to test and improve experimental designs, methods and analyses. In hindsight, the research questions still seem well-formulated, but the research designs are often clumsy and included unnecessary factors or conditions. For example, it would make more sense to study the roles of the auditory and visual modalities in a dedicated study rather than including that factor into every experiment. Also, usually synchronisation and continuation tasks are combined into one task rather than splitting them into separate conditions as was done in many of these studies. Likewise, a shift from traditional, static measures (average asynchrony) to more dynamic ones (windowed cross-correlations, \bar{R} entrainment measure) has proven helpful, as have better data visualisation techniques (e.g. rose histograms as part of the circular statistics approach).

Appendix B

Data collection

Initially, I set out to develop an experimental setup for rhythm and timing experiments which would include the social aspect that is missing from the usual tapping setups. I assumed that this would be a relatively easy addition, a simple quantitative change from one tapper to two tappers, but it turned out to be a qualitative step to a very different world. The main challenge was to develop data analysis methods and measures that would work in the situation where two people interact with each other, and that would capture relevant features of that interaction. Plugging in two MIDI drums instead of just one was very easy. Designing experimental procedures for cooperative tapping (as I decided to call these tapping tasks), was something in between. In this appendix, I discuss some of the issues related to data collection and research design in dyadic studies that came up during the project. I will focus on dealing with MIDI data and optical motion capture data.

B.1 Procedures

Developing procedures, tasks, and task instructions needed quite a lot of piloting. In the beginning, I was happy that there were many undergraduate music students willing to help me out in piloting and essentially making these pilot studies their own coursework projects (see more details of these pilot studies in appendix A). This allowed me to try out many different tasks and setup variations and manipulations. On one hand, this was extremely useful and informative, and I learned a lot in the process. But on the other hand it was a very time-consuming and fragmented way of working. None of these

small experiments was large or polished enough to be published as such, and they were all too different to be compiled into a coherent, larger study.

But the procedures got smoothened out, and I tested many different kinds of manipulations, and showed that it is feasible to:

- Manipulate stimuli so that only one participant receives a metronome, or that the two receive different metronomes. The metronome can of course also change tempo during the trial.
- Manipulate feedback channels so that the participants can either only hear each other, see each other, or both. Also, one can be made to hear the partner, while the other does not, creating a one-sided communication condition.
- Replace the partner's tapping with a pre-recorded computer-generated one.
- Manipulate feedback from participants' own tapping so that the feedback they receive is delayed or degraded in other ways.
- Vary tasks from simple synchronisation to syncopation, simple rhythms, alternating tapping (alternating bars), imitation etc.
- Introduce distractors and investigate their effects.
- Combine tapping tasks with other psychological measures in correlation designs.

These manipulations and their combinations can generate a lot of new experiments, many of which would complement the current literature on rhythm and timing.

Initially I assumed performing the basic task of synchronisation with a metronome would be easy even in a dyad. Ideally, all three onsets fall on the same time point, and having an extra tapper should not any difference. However, my very first pilot participants reported that they had difficulty choosing whether to stay with the metronome or tap with the other person. These two goals were in competition. This was exacerbated in the synchronisation-continuation studies, where in the absence of the metronome, following the "correct" tempo and negotiating a new one together were seen as very different goals.

Rather than just looking at these tasks as linear onset time optimisation, the participants see the metronome and the other tapper as intentional agents, and their intentions are seen as separate and, to some extent, conflicting. The metronome does not waver, but the other participant does exhibit some expressive timing or a tendency to speed up in the absence of the metronome. The joint intention is thus to be flexible, expressive, or to accelerate.

One aspect of procedure that has evolved over time are the instructions given to the participants. As participants have reported, keeping with the metronome and/or maintaining the original tempo can be in competition with maintaining a close synchrony with the other participant. Therefore it is necessary to tell the participants what to do re this conflict and which partner to prioritise. In the pilot studies participants usually prioritised the other person over the metronome. How much this is an effect of the mutuality, or just a question of social norms, where ignoring another human is more difficult than ignoring a computer is unclear, but some participants even drifted into a very different tempo while the metronome was still on, and did so in perfect synchrony among the participants, afterwards reporting that the trial went very well. The instruction that I ended up giving was thus that they need to keep the original tempo, but in collaboration with the other participant. This is saying that instead of choosing one, they would need to try to do both. But as the example above indicates, this is a conscious decision and an evolved procedure. Giving permission to ignore the metronome would mean that the participants would also ignore whatever manipulations the experimenter had made, such as tempo of the trial. On the other hand, ignoring the other participant, while it might be more difficult to do, would then defeat the whole purpose of the cooperative setting. In my experience, this instruction is easy to understand and most participants have no problems with it, they understand what it means after a couple of practice trials and can perform in a consistent manner. It would be interesting to see a proper experiment comparing different instructions, however.

B.2 Using MIDI

In my cooperative tapping setups, participants tap on a MIDI drum pad. In the first pilot studies, the drums had a dynamic range, so that the harder you

tap, the louder the sound. My reasoning was that this would help maintain the musical nature of the task, and keep the task more ecologically valid. This caused a number of problems, however. First, non-musicians (or musicians who were not percussionists) felt that having to control the loudness as well as timing was too difficult. To minimise the motor demands of the task, participants were tapping the drums with their fingers and not with the whole palm or a drum stick. Thus, loudness control would demand very fine motor control that was found to be too challenging without prior practice. Also, occasionally participants were competing for control and leadership in the interaction by increasing the loudness of their tapping. This would not only be bad for fingers (and the equipment) but also created an unnecessarily adversarial atmosphere in the experiment. Thus, the drum pads and the feedback sound generation were adjusted to be minimally responsive to changes in tapping force.

In my experiments, participants always got auditory feedback for their tapping. There are tapping studies with and without auditory feedback, but in the cooperative tapping studies, leaving it out would make the situation unnatural, as the partner needs to hear the output, anyway.

MIDI is a good option for music performance research, as the signal is much more compact than even poor quality audio. For rhythm and timing research, it's best feature is that you can access the onset times and durations of events (notes) without pre-processing. However, onset time extraction from audio is also getting relatively straight-forward thanks to advances in music information retrieval methods (Lartillot et al., 2008). It is especially easy from percussive sounds, so equipping a drum with a contact microphone or an acceleration sensor would work fairly well.

The MIDI data is read into MATLAB for further processing and statistical analysis. The MIDI Toolbox (Eerola & Toivainen, 2003) is very useful for these purposes. I have programmed a small toolkit of MATLAB functions that complement the MIDI toolbox for analysing the timing of events. For example, I have implemented the phase conversion and circular statistics functions as well as some pre-processing functions. FTAP, a tapping experiment toolkit for Linux, developed by Steven Finney (Finney, 2001), is a popular choice for rhythm production experiments on the Linux platform, and there are circular statistics extensions for R (e.g. Jammalamadaka & Sengupta, 2001), as well as very recently for MATLAB (Berens, 2009).

B.3 Movement data

The biggest limitation of MIDI and tapping data in general is that it is discrete and contains only the onset, velocity and duration of the events. This limitation was especially evident in the cooperative tapping study, where one participant used the inter-tap interval time to conduct the other participant. She subdivided the interval and made exaggeratedly high gestures to help her tapping partner to keep in tempo. Also, often the instruction is to start tapping as soon as you get the beat, which usually takes only 2 – 3 beats. However, most participants will wait for at least 4 beats (a full $\frac{4}{4}$ bar), sometimes even longer to commit, and start tapping, but you can see their feet tapping in time with the metronome from the second beat. So, the body is doing all kinds of things, and while the drum pad captures only a thin slice of them. Luckily, the motion capture technology is fast becoming available for music performance researchers.

There are many kinds of motion capture systems, and I will not attempt to make an exhaustive list or introduction of them all. Rather, I will just briefly introduce a couple of relevant methods. More details about how movement research works can be found in chapter 8 and e.g. in (Jensenius, 2008).

There are, in general, two kinds of motion capture systems. Both use markers at specific body points. In some systems the markers are active, in that they are devices that sense position and orientation (so-called 6DoF or six degrees of freedom measurement), and are attached to the bodies of the participants, and either wirelessly or with wires connect to the computer where the data is collected and stored. The Polhemus Fastrak¹ trackers that Shockley et al. used in their body sway study work this way (Shockley et al., 2003). The other type are optical motion capture systems using passive markers. In these systems (e.g. Vicon, Optitrack, Qualisys), the heart of the system are cameras that emit infrared light and record the reflections off reflective markers attached to the participants' bodies. The software is told where each of the cameras is positioned, and it then calculates the positions of the reflective markers up to 250 times per second. It is also possible to use rigid arrays of markers, which allows the system to make 6DoF measurements, similar to in the active sensor systems.

¹www.polhemus.com

The optical motion capture system produces a 3D representation of the movements with high temporal and spatial resolution. In the 8-camera system in Jyväskylä, the position of each marker is measured 120 times per second, at sub-millimetre accuracy. The 3D data can be visualised and thus the movement viewed from any angle, and of course quantitative analysis of the movement patterns of any individual body part can be carried out.

In a typical whole body marker setup, perhaps 30–40 markers are used so that all joints and body parts are recorded. As the position of every marker is recorded in three dimensions, this yields a relatively complex dataset. Usually for the analysis stage, the original marker set is reduced to a set of fewer markers that represent either the joints or body segments of the participants. For example, the wrist joint might be captured using two markers, one at the end of radius and the other at the end of ulna, and then for the analysis, these two are combined (averaged) to a virtual marker representing the wrist joint. A group of joint markers can then be converted into body segment models (Robertson, Caldwell, Hamill, Kamen, & Whittlesley, 2004). Also, the analysis often focuses on just a subset of markers, and to movement along just one spatial dimension. For example, in the body sway experiments, the anterior-posterior dimension of the torso markers was used. In the African dance study (see chapter 8), entrainment analysis was carried out looking at just the vertical dimension of the head markers. The MoCap Toolbox (Toiviainen & Burger, 2010) contains all the tools needed for the movement data analyses presented here.

The 3D position of a marker can only be recorded if it is visible to at least two cameras at the same time. Often as participants move, they occlude some of the markers occasionally, e.g. by covering their chest marker by their hands. This creates gaps in the data, which can be problematic, if the gaps are long or very frequent. Naturally, having more than one person in the capture volume makes these occlusions more likely, as the marker-camera lines of sight are now blocked by multiple bodies. For this reason, in the African choir study (reported in chapter 8), to be able to capture the movements of a relatively large group of people (8-10), each participant was wearing only three markers instead of the 30-40 markers in the full-body set.

Motion capture systems are expensive, but there are cheap alternatives for movement data collection. For example, most smart phones and the Nintendo Wiimotes come equipped with decent accelerometers, and can be

used to e.g. record step timings from walking participants, or the rhythm of dancing movements. Wiimotes can transmit the 3D acceleration data wirelessly over Bluetooth, and there is a free software WiiDataCapture ² that can record data from up to 8 Wiimotes simultaneously. The data can then be ported to MATLAB to be analysed with MoCap Toolbax. Capturing the vertical movement of the head (or upper torso) is often enough for entrainment analysis, as that movement component contains a good representation of the periodicity of the body movement (Toiviainen et al., 2010).

B.4 Audio and video

Collecting and analysing audio and video data is getting easier with the improvements of affordable high quality equipment and more sophisticated analysis software. Video analysis / annotation software such as Observer or ELAN ³ makes it easy to combine audio and video sources and make complex, multilayer annotations, view the video in slow motion and export annotation times for further analysis.

From the point of view of dyadic studies, the challenge is to capture what is relevant for interaction, and this often requires multiple camera angles and separate microphones for each participant. To investigate joint action, joint attention and gaze following are important factors, and thus the observer would need camera angles that can cover both participants and their gaze. In musical interaction, having an individual audio track that only has the feed from one participant is ideal, as this will let the audio analysis to be automatised and carried out quickly using, for example, the MIR ToolBox (Lartillot & Toiviainen, 2007). Having both participants in the same audio track or having significant leakage between the tracks will make for example a simple onset detection and onset synchrony analysis much more difficult.

²<http://bit.ly/WG1KRp>

³<http://tla.mpi.nl/tools/tla-tools/elan/>

Appendix C

Circular statistics

C.1 Illustration of the cross-over problem due to tempo drift

The crossover problem is common in tapping studies, and an example of it is demonstrated in figure C.1.

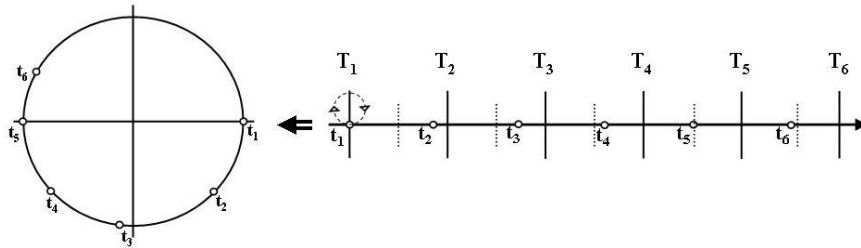


Figure C.1: Converting data from linear to circular domain

In this example, a tapper (onsets t_1, t_2, \dots, t_6) is tapping with a metronome (onsets T_1, T_2, \dots, T_6), but is speeding up and therefore drifting gradually ahead the beat. For a while, the taps are getting further and further away from the metronome onsets, until t_5 lands at the midpoint between two metronome taps T_4 and T_5 . The question is, is this tap still early or is it late? How about t_6 ? In the linear domain, this question would often need to be answered, for example to be able to calculate the onset time asynchrony of this tapper and the metronome. This is not necessary in the circular

domain, as the necessary metrics can be calculated from the distribution around the circle and there is no need to define which way around the circle the asynchrony (phase difference) is measured.

C.2 Additional phase conversion methods

C.2.1 Group sync: Abstracted beat as the referent

In group music-making, a steady tempo is usually maintained without the help of a metronome. In order to do this, the group abstracts a common beat that they all contribute to but no one explicitly or exclusively maintains. This abstracted beat could be estimated and used as a referent.

The mean onset time of the group could represent this abstracted, common beat. As no one participant is exclusively responsible for maintaining the common pulse, the average onset time could be a better option than using one performer as the referent. However, according to the models of Wing and Kristofferson (1973), Mates (1994), or Vorberg and Schulze (2002), the tappers' inter-onset intervals do not perfectly convey their sense of timing, but are "polluted" by clock and motor noises. Thus, to reduce the influence of any individual error or perturbation, a moving average of the group average onset times could serve as the local referent.

C.2.2 Polyrhythmic relationships

It is possible that our participants are not locked in one-to-one synchronisation, but instead have either a different harmonic or polyrhythmic relationship. In musical contexts, synchronisation patterns based on small integers ($1 : 1, 2 : 1, 3 : 1, 4 : 1, \text{etc.}$) are relatively common, and the harmonic ones ($n : 1$) are simple to deal with, both as a performer and from the point of view of data analysis. In these, one process has a period that is a small integer multiple of that of the other, or in the point of view of music theory, they are performing at different metrical levels, a beat of one being subdivided by the other.

In these cases, using the faster process as a referent gives a similar result than 1:1 entrainment would, as in both cases, the phase of the referent is expected to be zero at the time of the target tap. In the harmonic case the referent just makes multiple rounds around the unit circle during one period

of the target. However, in the harmonic or polyrhythmic cases and two processes, the selection of referent becomes crucial, as the expected phase values are different if the slower process is chosen as referent. For example, in the 2:1 case, when the slower process is the referent, the faster process is expected to be observed at both 0 phase difference and half way, at the anti-phase or π phase difference. This generates two clusters of vectors, effectively cancelling each other out, producing a mean resultant length \bar{R} (see below) of zero. Similar cancellation occurs with any harmonic n:1 relationship.

Polyrhythms are more difficult to perform (Deutsch, 1983; Kurtz & Lee, 2003; Povel, 1981), and the time relationships in these polyrhythms often approach simpler configurations, such as 2:1 (Fraisse, 1982; Repp, London, & Keller, 2012). In terms of analysis, a visual inspection of the data using rose histograms is a good way to analyse whether the data has polyrhythmic patterns. Or, you can take advantage of the fact that the small integer polyrhythms always have one cluster at the zero, as they meet every $n * m$ taps. So, with the proviso that the patterns may start from different taps, comparing the zero phase clusters formed by every $n * m^{th}$ onset, one can check which pattern might be most likely and then proceed in analysing it in more detail to see how well the correct pattern was performed.

C.3 Additional details on circular measures

C.3.1 Uniform probability plots

From rose histograms it might be difficult to see if the distribution is uniform (random) or not. Fisher (1993, 65-66) suggests that plotting the data as uniform probability plots will assist in visually inspecting the distribution. The process of producing this plot starts from rank-ordering the phase angles from the smallest to the largest, and then dividing them by 2π , plotting these against the index of their position in the sequence (see figure C.2 and Fisher's equations 4.5 and 4.6). If the data is uniformly distributed, it will fall along a straight line in a 45 degree angle, passing through the origin.

C.3.2 Direction and spread of sample

The measure of concentration of circular distribution, \bar{R} , is the mean resultant length of the sum vector, obtained by summing up all vectors making up

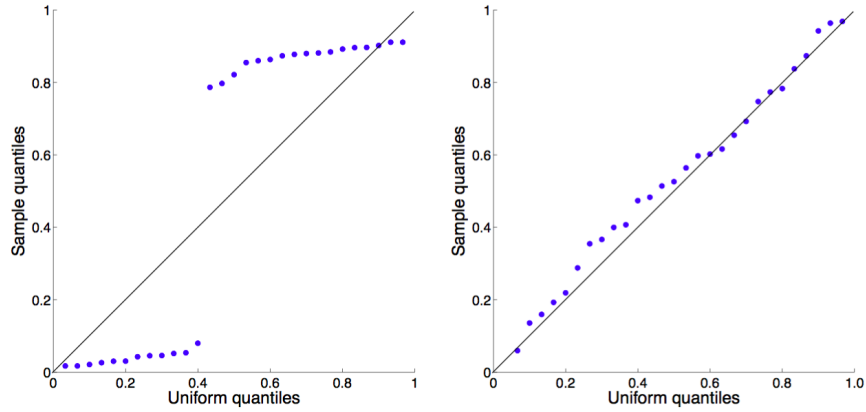


Figure C.2: Uniform distribution. Data on the left panel is from a unimodal distribution (relative phase of two entrained tappers), while the data on the right panel comes from a uniform distribution (relative phase of one tapper and an unrelated metronome). Data is mean-normalised and rank-ordered and plotted against the rank. Uniformly distributed data would follow the diagonal (black line).

the circular distribution. For this calculation, each onset time is converted to unit phase vector (from origin, length 1), using one of the methods presented in section 5.1.5. This measure \bar{R} can be used as entrainment measure or stability measure. This sum vector has a direction, θ , which is the mean direction of the distribution, which in the case of entrainment analysis represents the average asynchrony between performers. While these two measures often provide sufficient information about the trial, there are other measures that may be useful.

The reciprocal of \bar{R} , or the quantity $V = 1 - \bar{R}$ is known as *sample circular variance*. The range of circular variance is naturally from 0 to 1 just like that of the \bar{R} , and the smaller the V the more concentrated the distribution. This can be a very useful quantity, as conceptually it is easy to link it with variance in linear domain. The choice between v and \bar{R} is a matter of whether it makes more sense to “award” higher scores for higher concentration/stability, or for higher values of variance. In the linear domain, standard deviation is the square root of variance, but in circular domain, the *sample circular standard deviation* is not \sqrt{V} , but rather it is defined as

$$v = -2\log(1 - V)^{\frac{1}{2}} \quad (\text{C.1})$$

According to Fisher (1993, p. 33), this has to do with how population parameters are defined in circular domain. But Fisher offers a more handy way of approximating v , and again, these only work if the data is in unimodal distribution with relatively low variance / high \bar{R} :

$$V = [2(1 - \bar{R})]^{\frac{1}{2}}, \text{ for large values of } \bar{R} \quad (\text{C.2})$$

This emphasises again the importance of plotting the data before running even descriptive analyses.

Another useful measure of the spread of the distribution is *sample circular dispersion* $\hat{\delta}$, which is defined as

$$\hat{\delta} = \frac{1 - \hat{\rho}_2}{2\bar{R}^2} \quad (\text{C.3})$$

where

$$\rho_2 = m_2 = \left(\frac{1}{n}\right) \sum_{i=1}^n \cos 2(\theta_i - \bar{\theta}) \quad (\text{C.4})$$

The $\hat{\delta}$ is an important measure for the analytical statistics, as it is used as a confidence measure in comparing mean directions. It ranges from zero of perfect concentration to infinity for uniform distribution.

Appendix D

Cross-recurrence parameters

D.1 Parameter selection

From a practical perspective, the most difficult part of cross-recurrence analysis is determining optimal values of the parameters. In the field of nonlinear dynamics, various methods for estimating optimal delays and embedding dimensions have been developed, but these often require knowledge about the underlying dynamics of the system. This can be tricky when studying interpersonal processes. Marwan et al. (2007) provide a summary of the different methodical details of cross-recurrence quantification (CRQ), especially the different ways in which recurrence and cross-recurrence can be defined. The Cross Recurrence Plot Toolbox for MATLAB (Marwan, 2012), takes the fuss out of plotting recurrences, and the graphical user interface is helpful.

In terms of what the data can be, CRQ is rather forgiving. For the CRQ analysis to work, more or less the only limitation is that the two data sets need to roughly be on the same scale. This is usually easily achieved by normalising the variables. In the example of dyadic improvisation and conversation, we followed examples found in the literature (Shockley, Baker, Richardson, & Fowler, 2007; Shockley et al., 2003), we used the displacement of the chest markers in the anterior-posterior dimension, and normalised the data by subtracting the time series values from its mean. This brought the values of the two time series to approximately the same range of values. One advantage of CRQ is that the time series do not need to be stationary (their mean and standard deviation can change over time¹), and

¹Although in non-stationary cases, different normalisation methods are needed.

the analysis works relatively well with noisy data, as well, which means that the data does not necessarily need to be extensively filtered or de-trended before the analysis.

The difficulty in using CRQ is in the selection of the parameters of the analysis. The results can change wildly depending on the parameters, and there is no automatic or fool-proof way to set these parameters. There are however, ways to determine “safe” ranges for these parameters. The main parameters that you will need to set are a) the delay at which the embedding takes place; b) the number of embedding dimensions; c) the threshold or the distance in phase space where the two processes are considered to have recurred.

D.1.1 Delay

Delay, or τ is a parameter used in reconstruction of the phase trajectory from one time-series variable. Selecting delay parameters requires a consideration of sampling rate of the data, and relevant time scales of the underlying processes. For the tapping data, a suitable delay is 1, as each data point represents an event; this data is discrete. The movement data, which is quasi-continuous, is recorded at 120 Hz temporal resolution, while individual gestures or movements would take anything from a few tenths of a second to a couple of seconds. Thus an embedding delay of just 1 sample would be probably too short, and on the other hand, embedding delay of 120 samples, representing a whole second, would probably be too long. One method offered for estimation of the delay parameter is to calculate the mutual information between the time series and its delayed version, using a range of time delays (Cao, 1997; Fraser & Swinney, 1986). The delay value that has the minimum of mutual information is a good candidate for delay parameter. A practical range might be somewhere between 20 and 60 samples (representing $\frac{1}{6} - \frac{1}{2}$ seconds of movement).

D.1.2 Embedding dimension

The choice of embedding dimension (d) depends on the dimensionality of the underlying system. There are a number of heuristics for choosing the embedding dimensions, one of which is that one should choose a high value first and then reduce it, if necessary, computation-wise. The problem of using too

high a dimension is that the noise in the system has infinite dimensionality while the actual system output has only a limited number of dimensions. Thus, using a large number of dimensions will mean that the noise is captured from all the dimensions while only the first N dimensions will provide information about the system's behaviour. Thus using too many dimensions means you get more noise but not more signal. On the other hand, using too few dimensions means that the phase trajectory will remain entangled resulting in false recurrences appearing in the recurrence plots.

The phase space of the system is reconstructed from a single variable by using that variable and time-delayed copies of it. An example of this can be seen in figure 5.7, where the inter-onset intervals of one tapping trial are plotted against two time-delayed versions of it; the version delayed by one tap and the version delayed by two taps. In this case, then, the number of embedding dimensions is three and the delay is 1. When rotated to the viewing angle you see in the picture, a limit cycle of sorts seems to emerge in the phase space, with the phase trajectory forming a triangular periodical loop. In contrast, the same data plotted using only two dimensions is a much more messy bundle, the trajectory does cross itself many times (all these would be recurrence points), but no such structure is apparent.

Three dimensions are easy to visualise, but that does not mean the system's behaviour could not have many more dimensions. And just as in the figure 5.7, adding one more dimension helps to “disentangle” the phase trajectory, this disentanglement should be done with the data before trying to calculate the recurrences. The fact that each added dimension helps in this disentanglement has given rise to a very good algorithm for estimating the correct number of dimensions in recurrence analysis. The algorithm is called *false nearest neighbours* or *fnn* for short (Cao, 1997). This algorithm starts with a small number of dimensions, and checks which data points seem to recur. Then, one more embedding dimension is added, and if the two points no longer are next to each other, they can be deemed as false near neighbours, as they only seemed to be near each other in the lower dimensional projection, but as another dimension was added, we could see that there was actually distance between them in this new dimension (as in many cases in 5.7). Continuing to higher dimensions, at some point the number of *fnn*'s drops to zero, and this can be considered the point where the phase space trajectory is fully disentangled.

D.1.3 Threshold

Two timeseries are considered jointly recurring, when the trajectory of one is closer to the other than a certain threshold distance ϵ . The choice of this threshold value, the distance separating the two trajectories for recurrence to be considered to have happened, naturally has a large influence on how much recurrence there is in the system. If the distance between the trajectories can be very large, then almost at any point there is cross-recurrence. On the other hand, if the trajectories need to be extremely close for it to count as cross-recurrence, recurrences will be few and far in between. A heuristic suggested by Webber (see Shockley, 2005, 166) is that one should aim for a relatively sparse cross-recurrence matrix, so that the recurrence rate is approximately 1–3%.

In addition to selecting a constant threshold distance, it is also possible to do as Shockley et al. (2003) did for their analysis, and choose a threshold that adjusts separately for each trial, depending on the mean distance between the trajectories. This serves as a kind of scaling between the trials, helping to take into account that the size of the phase space can vary between trials.

D.2 Example: parameter selection in improvisation trial

In this section, I will detail the parameter estimation process by using dyadic shaker improvisation example discussed in section 5.2.3. The data was position data from optical motion capture, measured at 120 frames per second. Head, chest, root and hand markers were selected for analysis.

First, to estimate appropriate embedding dimensions (d), we used the false nearest neighbours algorithm. Using segments from a number of different trials, the highest dimension where fnn's were found was 7, although in one data set (with a marker flicker problem that resulted in discarding the data altogether) there was a non-zero fnn score for dimensions 7, 9 and 17, while 6th dimension already had a zero fnn. In most cases, the first non-zero dimension was the 4th or 5th one. Given that according to Kennel, Brown, and Abarbanel (1992), this algorithm tends to slightly underestimate the number of dimensions needed for phase space reconstructions, and as Shockley et al. used $d = 10$, we decided to settle for $d = 8$.

In terms of delay, phenomenally we expected the appropriate delay to be in the range of few tenths of a second to one second. After testing out a number of values from that range, and plotting the mutual information over time delay for the different trials, we settled for delay of 40 samples, representing $\frac{1}{3}$ of a second.

In (Thiel et al., 2002) it was suggested that observational noise can cause large errors in estimating the determinism in the CRQ, and they also suggested that the threshold ϵ should be set at 5 times the standard deviation of the gaussian noise in the observations. Instead of trying to measure the systemic noise, we followed the heuristic used by Shockley et al. (2003). In their study, the threshold was set so that it yielded a low but relatively similar recurrence across trials. We wanted to eliminate possible floor and ceiling effects in either recurrence rate or determinism, and aim for recurrence rates between 1–5%. Shockley et al. (2003, 329) suggest that one should calculate the recurrence rates using a range of radius values and then plot the results in log-log coordinates. The suitable range for ϵ should be found in a region where this plot is roughly linear. Based on doing this for radius values of 0 – 1 (Euclidean distance), the value of $\epsilon = 0.8$ was selected. Some results from this calculation are presented in 5.2.3.

Appendix E

Publications

E.1 Refereed articles

Himberg, T. & Thompson, M.R. (2011). Learning and synchronising dance movements in South African songs - Cross-cultural motion-capture study. *Dance Research*, 29(2), 305–328.

London, J., Himberg, T. & Cross, I. (2009). The Effect of Structural and Performance Factors in the Perception of Anacruses. *Music Perception*, 27(2), 103-120.

Eerola, T., Himberg, T., Toiviainen, P. & Louhivuori, J. (2006). Perceived Complexity of Western and African Folk Melodies by Western and African Listeners. *Psychology of Music*, 34(3), 337-371.

E.2 Conference proceedings papers

Himberg, T. (2012). Moleminpuolinen mukautuminen: Tahdistumisen ja musiikillisen vuorovaikutuksen tutkimus. (Mutual adaptation: Research on entrainment and musical interaction). In Himberg, T., Vuoskoski, J., and Eerola, T., editors, *Proceedings of the 16th Annual Symposium for Music Scholars in Finland, Jyväskylä, Finland*. University of Jyväskylä.

Spiro, N. & Himberg, T. (2012). Musicians and Non-musicians Adapting to Tempo Differences in Cooperative Tapping Tasks. In E. Cambouropoulos, C. Tsourgas, P. Mavromatis, and K. Pasteriadis. *Proceedings of the*

12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece.

Himberg, T., Braithwaite, L., Snape, J. & Spiro, N. (2011). Negotiating Chaos - A View to Entrainment. In *Actas del X Encuentro para las Ciencias Cognitivas de la Música, SACCoM. Buenos Aires, Argentina.*

Himberg, T. & Thompson, M.R. (2010). Dyadic entrainment and interaction in African dance. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.). *Proceedings of the 11th International Conference for Music Perception and Cognition. Seattle, US: University of Washington.*

Himberg, T. & Thompson, M.R. (2009). Group Synchronization of Coordinated Movements in a Cross-cultural Choir Workshop. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (eds.). *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music, (ESCOM). Jyväskylä, Finland: University of Jyväskylä.*

Himberg, T. (2008). Dyadic Entrainment and Interaction in African Dance. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima & M. Tsuzaki, (Eds.). *Proceedings of the 10th International Conference on Music Perception and Cognition, (ICMPC). Sapporo, Japan.*

E.3 Conference proceedings (as editor)

Himberg, T., Vuoskoski, J., & Eerola, T. (eds.) (2012). *Proceedings of the 16th Annual Symposium for Music Scholars in Finland, Jyväskylä, Finland. University of Jyväskylä.*

Louhivuori, J., Eerola, T., Saarikallio, S., Himberg, T. & Eerola, P.-S. (eds.) (2009). *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music, (ESCOM). Jyväskylä: University of Jyväskylä.*

E.4 Invited and keynote lectures

Himberg, T. (2010). Interaction in African Dance – A Cross-cultural study. New Perspectives on Joint Action and Task Sharing – workshop. Max Planck Institute for Human Cognitive and Brain Sciences. Leipzig, Germany. Invited lecture.

Himberg, T. (2008). Co-operative Tapping: Musical and Social Interaction. In Symposium “Rhythmic Coordination in Dyads”. Max Planck Institute for Human Cognitive and Brain Sciences. Leipzig, Germany. Invited lecture.

E.5 Conference presentations and posters

Gill, S., Thompson, M.R. & Himberg, T. (2012). Body Rhythmic Entrainment and Pragmatics in Musical and Linguistic Improvisation Tasks. 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece. Spoken presentation.

Himberg, T. (2012). Molemmipuolinen mukautuminen: Tahdistumisen ja musiikillisen vuorovaikutuksen tutkimus. (Mutual adaptation: Researction). 16th Annual Symposium for Music Scholars in Finland, Jyväskylä, Finland. Spoken presentation.

Himberg, T. & Spiro, N. (2012). Beating to Each Other’s Drum: Towards a comprehensive understanding of pairwise rhythmic interaction. 7th Nordic Music Therapy Congress, Jyväskylä, Finland. Spoken presentation.

Hoadley, R., Aaron, S., Cross, I., Gill, S., Himberg, T., Odell-Miller, H. & Toulson, R. (2012). Touching Sound: Technologies for mediated interaction in music therapy. 7th Nordic Music Therapy Congress, Jyväskylä, Finland. Spoken presentation (delivered by Dr. Hoadley).

Spiro, N. & Himberg, T. (2012). Empathy and Resisting Entrainment: Mapping the dimensions of pairwise rhythmic interaction. 16th Annual Symposium for Music Scholars in Finland, Jyväskylä, Finland. Spoken presentation.

- Spiro, N. & Himberg, T. (2012). Musicians and Non-musicians Adapting to Tempo Differences in Cooperative Tapping Tasks. 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece. Poster.
- Gill, S.P., Himberg, T. & Thompson, M.R. (2011). Pragmatics of Bodily Rhythm. BPS3: The Music of Language, The Language of Music: Communicative Interaction. Middlesex University. Poster.
- Himberg, T., Braithwaite, L., Snape, J. & Spiro, N. (2011). Negotiating Chaos - A View to Entrainment. X Encuentro para las Ciencias Cognitivas de la Música, SACCoM. Spoken presentation (delivered by Dr. Spiro).
- Himberg, T. (2011). Interacting with Non-responsive and Responsive Tapping Partners. 13th International Rhythm Perception and Production Workshop. Leipzig, Germany. Poster presentation.
- Himberg, T., & Thompson, M.R. (2010). Dyadic entrainment and interaction in African dance. In 11th International Conference for Music Perception and Cognition. Seattle, US. Poster.
- Himberg, T. & Thompson, M.R. (2009). Group synchronization of coordinated movements in a cross-cultural choir workshop. In 7th Triennial Conference of the European Society for the Cognitive Sciences of Music, (ESCOM). Jyväskylä, Finland. Spoken presentation.
- Himberg, T. (2009). Analysing Group Entrainment. In Entrainment Network. Milton Keynes, UK. Spoken presentation.
- Himberg, T. (2008). Personality Types and Musical Interaction. In IMHSD Conference: Dance, Timing and Musical Gesture. Edinburgh, UK. Poster.
- Himberg, T. (2008). Measuring Co-operability in Tapping Dyads. In 10th international conference for music perception and cognition, (ICMPC). Sapporo, Japan. Poster.

- Himberg, T. & Thompson, M.R. (2010). Dyadic entrainment and interaction in African dance. In 11th International Conference for Music Perception and Cognition. Seattle, US. Poster.
- Himberg, T. (2006). Co-operative tapping and collective time-keeping – differences of timing accuracy in duet performance with human or computer partner. In ICMPC9 (International Conference for Music Perception and Cognition), Bologna, Italy (24.8.2006). Spoken presentation.
- Himberg, T. (2006). Cooperative tapping with human and computer partners – investigating shared intentions and interaction in musical time. In Rhythm, Time & Temporal Organisation, Edinburgh, Scotland (4.6.2006). Spoken presentation.
- Himberg, T. (2005). Cognitive Foundations of Musical Interaction – Exploring Cooperative Tapping. In 10th International Rhythm Perception and Production Workshop (RPPW10), Bilzen, Belgium (5.6.2005). Spoken presentation.
- Himberg, T. (2004). Exploring the Cognitive Foundations of Musical Interaction. In Nordic Musicology Conference (NMK2004), Helsinki, Finland (13.8.2004). Spoken presentation.
- Himberg, T. & Cross, I. (2004). Interaction in Musical Time. In ICMPC8 (International Conference for Music Perception and Cognition), Evanston, IL (6.8.2004). Spoken presentation.
- Himberg, T. (2004). The Cognitive Foundations of Musical Interaction. In SEMPRES Conference “Musical Collaboration” Milton Keynes, UK (24.4.2004). Spoken presentation.

E.6 Teaching materials

- Lähdesmäki, T., Hurme, P., Koskimaa, R., Mikkola, L. & Himberg, T. (2010). Menetelmäpolkuja humanisteille. University of Jyväskylä, Faculty of Humanities. <http://www.jyu.fi/mehu> (Mapping Research Methods; English version <http://bit.ly/11g4N11>)