Title

Rhythms at Small and Large Scales: the Neural Mechanisms of Rhythm Perception and the Recurrence Dynamics of Large Group Interaction

Author

Proksch, Shannon

Publication Date

2022

UNIVERSITY OF CALIFORNIA, MERCED

Rhythms at Small and Large Scales: the Neural Mechanisms of Rhythm Perception and the Recurrence Dynamics of Large Group Interaction

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Cognitive and Information Sciences

by

Shannon Proksch

Committee in charge:

Professor Ramesh Balasubramaniam, Ph.D., Chair
Professor Kristina Backer, Ph.D.
Professor Christopher Kello, Ph.D.

2022

Rhythms at Small and Large Scales: the Neural Mechanisms of Rhythm Perception and the Recurrence Dynamics of Large Group Interaction

The dissertation of Shannon Proksch is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Professor Christopher Kello, Ph.D.

Professor Kristina Backer, Ph.D.

Professor Ramesh Balasubramaniam, Ph.D., Chair

University of California, Merced
2022

To my parents, Brent and Anne, for supporting me no matter how ambitious
(and annoying) I could be.

To my partner, Byron, for keeping me grounded as I traveled across
continents on this arduous academic journey.

To my brother, Brian, for always making sure to visit.

Thor, thanks for coming along.

Contents

List of Figures

List of Tables

Acknowledgments

External Funding

This material is based upon work supported by the National Science Foundation National Research Traineeships in Intelligent Adaptive Systems under Grant No. DGE-1633722 and in Interdisciplinary Computational Graduate Education Grant No. DGE-1633429. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Internal Funding

This material is based upon work supported by the University of California, Merced Graduate Dean's Recruitment Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of California, Merced.

Collaborators / Co-authors

Majerle Reeves — Applied Mathematics, UC Merced
Alexandria Pabst — Accenture
Daniel Comstock — Center for Mind and Brain, UC Davis
Butovens Médé — Cognitive & Information Sciences, UC Merced
Jessica M. Ross — Pychiatry and Behavioral Sciences, Stanford Medical School
Michael Spivey — Cognitive & Information Sciences, UC Merced
Kent Gee — Physics and Astronomy, Brigham Young University
Mark Transtrom — Physics and Astronomy, Brigham Young University

Permission to use copyrighted material in Chapters 1 and 4 was automatically granted by Creative Commons Attribution Licences (CCBY 4.0). Permission to use copyrighted material in Chapter 3 was obtained by Brill Publishers. Permission to use unpublished material in Chapter 5 was obtained from all co-authors.

Personal Acknowledgements

To my labmates, Alexandria Pabst, Daniel Comstock, Butovens Médé, Jessica Ross, Chelsea Gordon, Tim Shea, Jaskanwaljeet Kaur, Sam Carey, and Tadhg Waddington – thank you for all of your guidance, support, and friendship. To my research partner and friend, Majerle Reeves, for all of our insightful conversations and collaborations, and for keeping me grounded when I was focused on the idealistic. To my committee, Kristina Backer, Chris Kello, and my advisor Ramesh Balasubramaniam – thank you for the opportunities and doors you opened, for your guidance, and your trust.

To my dedicated research assistants, Arianna Malakis, Anand Sivaram, Yash Sharma, Riley Fette, Thomas Dvorchkin, and Shani Campbell – thank you for diligent data collection work, for lending your computational skills, for your amazing spirit during the pandemic, and for your constant curiosity. This work wouldn't have been possible without your help!

To all the faculty and graduate students of UCM CogSci, especially Sara Schneider, Camila Alviar, and Ayme Tomson, thank you for working to build the kind of community that is supportive of everyone in both their life and research endeavors.

To my undergraduate mentors at Texas A&M University - Corpus Christi, Mary Thornton, Miguel Moreno, Evelyn McCarty, and Don Berkich – thank you for opening up the worlds of music, philosophy, and cognitive science to me. Thank you for holding me to the highest standard, whether in performance or in research, and encouraging me when I decided I wanted to try my hand at this science thing, without leaving music behind.

To the dear friends I made during my time at the University of Edinburgh – Kate Nave, Kate Webb, Matt Sims, Danaja Rutar, George Deane, Matthew Crosby, and Viky Neacsu, thank you to each of you for the stimulating conversations, the travels, and overall enthusiasm and excitement for the philosophical, experimental, and computational world of human cognition.

Finally, to my family, for supporting me no matter how far away I traveled or how many more years of school I sought.

[This page intentionally blank for formatting purposes.]

Curriculum Vitae

Shannon Proksch

www.shannonproksch.com

EDUCATION

PhD Cognitive and Information Sciences **est. grad. 2022**
University of California, Merced
Advisor: Prof. Ramesh Balasubramaniam

MSc, Mind, Language and Embodied Cognition **2017**
with Distinction
The University of Edinburgh

BA, Music, minors Philosophy & Psychology **2015**
Summa Cum Laude, Honors Program Graduate with Thesis Defense
Texas A&M University – Corpus Christi

Certificate, Teaching English to Speakers of Other Languages **2015**
Texas A&M University – Corpus Christi

Seven-week French Language Immersion Program **2013**
Middlebury College, Vermont

ADDITIONAL TRAINING

Institute for Evidence-Based Teaching Practices
Center for Engaged Teaching & Learning, UC Merced, California 2021
Neuromatch Academy Computational Neuroscience Summer School 2020
American Psychological Association Advanced Training Institute:
Nonlinear Methods for Psychological Science, University of Cincinnati, Ohio 2019
Spring School Language and Music in Cognition:
Integrated Approaches to Cognitive Systems, Universität zu Köln, Cologne, Germany 2019
Linear Models and Mixed Models with R, University of Edinburgh (Instructor: Bodo Winter) 2017

AWARDS & SCHOLARSHIPS

Cognitive & Information Sciences Graduate Student Spotlight Award Fall 2021
for outstanding contributions to scholarship and service, *UC Merced*

NSF Research Traineeship (NRT) Fellowship-
Intelligent Adaptive Systems *UC Merced* Fall 2021-Summer 2022

NSF Research Traineeship (NRT) Fellowship-
Intelligent Adaptive Systems *UC Merced* Fall 2019-Summer 2020

NRT Interdisciplinary Computational Graduate Education Program *UC Merced* Spring 2019

Travel grant from the American Psychological Association Summer 2019
for the Advanced Training Institute: Non-Linear Methods for Psychological Science

Graduate Dean's Recruitment Fellowship *UC Merced* Fall 2018-Summer 2019

Highly Commended Dissertation Prize *University of Edinburgh* October 2017

Conference Travel Funding Award (*University of Edinburgh*; *for presentation at the 6th Dubrovnik Conference on the Philosophy of Art*) April 2017

PUBLICATIONS & IN PRESS

Proksch, S., Reeves, M., Spivey, M., Balasubramaniam, R. (*in press*) Measuring acoustic social worlds: reflections on a study of multiagent human interaction. In S. Besser, F. Lysen, N. Geode (Eds.) *Worlding the Brain*, forthcoming with Brill in 2022.

Pabst, A., **Proksch, S.,** Médé, B., Comstock, D., Ross, J. M., & Balasubramaniam, R. (2022). A systematic review of the efficacy of intermittent theta burst stimulation (iTBS) on cognitive enhancement. *Neuroscience and Biobehavioral Reviews.* Preprint doi: 10.31234/osf.io/me6n4 Publication doi: 10.1016/j.neubiorev.2022.104587

Proksch, S., Reeves, M., Balasubramaniam, R., & Spivey, M. (2022). Coordination Dynamics of Multi-Agent Human Interaction in a Musical Ensemble. *Scientific Reports. 12,* 421.
Preprint doi: 10.31234/osf.io/nz4g2
Publication doi: 10.1038/s41598-021-04463-6

Moser, C., Ackerman, J., Dayer, A., **Proksch, S.,** Smaldino, P.E. (2021) Why don't cockatoos have war songs? [commentary on Mehr et al. and Savage et al.] *Behavioral and Brain Sciences*. Preprint doi: 10.31234/osf.io/j3nd6 Publication doi: 10.1017/S0140525X20001223

Holbrook, C., Iacoboni, M., Gordon, C., **Proksch, S.,** Makhfi, H., & Balasubramaniam, R. (2021). Posterior Medial Frontal Cortex Regulates Sympathy: A TMS Study. *Social Neuroscience.* Preprint doi: 10.31234/osf.io/hduac Publication doi: 10.1080/17470919.2021.1980097

Pouw, W., **Proksch, S**., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R., Wiggins, G. (2021). Multilevel interactions in multimodal communications. *Philosophical Transactions of the Royal Society B: Biological Sciences.* Preprint doi: 10.31219/osf.io/psmhn
Publication doi: 10.1098/rstb.2020.0334

Proksch, S. Comstock, D. C., Médé, B., Pabst, A., Balasubramaniam, R. (2020). Motor and predictive processes in auditory beat and rhythm perception. *Frontiers in Human Neuroscience, Cognitive Neuroscience.* 14:578546. doi: 10.3389/fnhum.2020.578546

Holbrook, C., Iacoboni, M., Gordon, C., **Proksch, S.**, Balasubramaniam, R. (2020). Posterior medial frontal cortex mediates threat-enhanced religious belief: A replication an extension study. *Social Cognitive and Affective Neuroscience,* nsaa153. doi: 10.1093/scan/nsaa153

Proksch, S. (2017). Interoceptive inference and emotion in music: Integrating the neurofunctional 'Quartet Theory of Emotion' with predictive processing in music-related emotional experience. *Journal of Cognition and Neuroethics,* 5(1): 101-125.

Proksch, S. (2017). *Integrating the neurofunctional 'Quartet Theory of Emotion' with predictive processing in music-related emotional experience.* (Unpublished master's dissertation). The University of Edinburgh, Edinburgh, Scotland.

Proksch, S. (2015). *Music and language: Evolutionary and cognitive perspectives.* (Unpublished BA Honors Thesis). Texas A&M University-Corpus Christi, Corpus Christi, Texas.

Proksch, S. (2014). *Music and language in the mind: Intention, structure, and perception.* (Unpublished BA Thesis). Texas A&M University-Corpus Christi, Corpus Christi, Texas.

PREPRINTS & SUBMISSIONS

Ross, J. M.*, **Proksch, S.*,** Iversen, J., Balasubramaniam, R. (*submitted*). Contributions of left and right parietal cortex for musical beat perception: replication and extension. **These authors contributed equally*

Rahimpour, A., **Proksch, S.,** Lanka, P., Pollonini, L., Balasubramaniam, R., Bortfeld, H. (*submitted*). Multiple levels of contextual influence on action-based timing behavior and cortical activation.

CONFERENCE PRESENTATIONS

Proksch, S., (October 2021). Acoustic social worlds: from Markov blankets to interpersonal synergies. Talk presented at *Cognitio 2021: Active inference and collective intelligence.* Montréal, Canada.

Proksch, S., Reeves, M., Spivey, M., Balasubramaniam, R. (August 2021). Coordination dynamics of multi-agent human interaction–recurrence quantification analysis of a musical performance. Poster presented at the *16th International Conference on Music Perception and Cognition and 11th triennial conference of the European Society for the Cognitive Sciences of Music.* Sheffield, United Kingdom.

Proksch, S., (November 2019). Synchrony and the Markov Blankets of Multi-Agent Human Interaction. Talk presented at the *Timescales of Life and Mind: Interdisciplinary Conference on Timescales as a Methodological Approach to Predictive Processing and the Free Energy Principle,* Edinburgh, Scotland.

Ross, J. M., **Proksch, S.,** Iversen, J.R., Balasubramaniam, R. (August 2019). Down-regulation of left posterior parietal cortex impairs musical phase shift detection in subjects with good perceptual acuity for phase timing. Poster presented at the *Biennial Meeting of the Society for Music Perception and Cognition*. New York City, USA.

Ross, J. M., **Proksch, S.,** Iversen, J.R., Balasubramaniam, R. (July 2019). Down-regulation of left posterior parietal cortex impairs musical phase shift detection in subjects with good perceptual acuity for phase timing. Poster presented at *Synchrony and Rhythmic Interaction: From Neurons to Ecology Lorentz Workshop,* Leiden, Netherlands.

Proksch, S., Booth, L., Dayer, A., Reeves, M., Spivey, M. (June 2019). Online and In Sync? Bodily synchrony & interpersonal closeness in audiovisual communication. Poster presented at the *American Psychological Association Advanced Training Institute on Nonlinear Methods for Psychological Science*, Cincinnati, Ohio.

Ross, J. M., **Proksch, S.,** Iversen, J.R., Balasubramaniam, R. (March 2019). Hemispheric differences in parietal contributions to auditory beat perception. Poster presented at *Cognitive Neuroscience Society 2019 Annual Meeting,* San Francisco, California.

Proksch, S., Ross, J. M., Balasubramaniam, R. (February 2019). Hemispheric differences in parietal contributions to auditory beat perception. Poster presented at *Spring School 2019 Language and Music in Cognition: Integrated Approaches to Cognitive Systems*, Cologne, Germany.

Proksch, S. (November 2017). Interoceptive inference and emotion in music: Integrating the neurofunctional 'Quartet Theory of Emotion' with predictive processing in music-related emotional experience. Poster presented at *Expecting Ourselves: Prediction, Action, and the Embodied Mind*, Edinburgh, United Kingdom, November 2017.

Proksch, S. (November 2017). Grounding musical entrainment and social cohesion in sensorimotor contingencies. Talk presented at the *Worlding the Brain Conference*, Amsterdam, Netherlands.

Proksch, S. (August 2017). Interoceptive inference and emotion in music: Integrating the neurofunctional 'Quartet Theory of Emotion' with predictive processing in music-related emotional experience. Talk presented at the *Mind and Brain Annual Conference*, Flint, MI.

Novelli, N. & **Proksch, S.** (April 2017). The limitations of music making machines. Talk presented at the *6th Dubrovnik Conference on the Philosophy of Art*, Dubrovnik, Croatia.

OTHER PRESENTATIONS

Frontiers of Science Institute – University of Northern Colorado June 2021
Presented my research to high school students interested in pursuing STEM research in higher education; Discussed transitioning from an arts and humanities background into STEM-oriented fields
Topic: Music, Movement, and the Brain

Research Seminar in Systematic Musicology – Universität Graz June 2021
Topic: Coordination dynamics of multi-agent interaction within a musical ensemble

Achievement Rewards for College Scientists Foundation March 2021
Presented to the ARCS Northern California Chapter
Topic: Overview of my music/brain work, & graduate student life at UC Merced

Guest Lecturer – University of North Carolina, Charlotte January 2021
Presented to undergraduate music students of the UNC-Charlotte Wind Bands
Lecture Topic: Neuroscience, Music, & Emotion

Workshop Leader – Professional Development Event October 2020
Led an interactive workshop for the Cognitive & Information Sciences Graduate Student Group
Workshop Topic: Giving Effective Presentations at Online Conferences

Guest Lecturer – COGS 001, Intro to Cognitive Science July 2020
Lecture Topic: Neuroscience and Music

RadioBio GradStory March 2020
Interviewed by RadioBio, a graduate student run podcast, during UC Merced's Research Week
Interview topic: Rhythms in Brains and Interacting People.

Cognitive Science Student Association Symposium December 2019
Outreach event sharing graduate student cognitive science research with undergraduate students
Talk title: On Music & Artificial Intelligence: The limitations of music-making machines

Edinburgh Music Psychology Research Reading Group Spring 2017
Organized and led discussion for three reading group sessions on:
Introduction to music and predictive processing
Emotion in music and predictive processing
Rhythm in music and predictive processing

TEACHING EXPERIENCE

Instructor – Introduction to Neuroscience June 2022
Frontiers of Science Institute – University of Northern Colorado
Instructor of Record – PHIL / COGS 110, Philosophy of Cognitive Science Summer 2021
University of California, Merced
Teaching Assistant – PHIL / COGS 110, Philosophy of Cognitive Science Spring 2021
University of California, Merced
Teaching Assistant – COGS 140, Perception and Action Fall 2020
University of California, Merced
English Second Language Instructor Nov 2016—May 2018
E-Joy English Online Language School, Beijing, China
Elementary Music and English Second Language Instructor Feb 2016—June 2016
Ameri-Can International Academy,
CIBT Education Group, Weifang, China
Grader/Supplemental Instructor- Honors Elementary Formal Logic Spring 2015
Texas A&M University- Corpus Christi
Reader/Grader- Music Appreciation Fall 2014
Texas A&M University- Corpus Christi

Counselor/Music Theory Instructor- Islander Band Camp 2013—2015 (summers)
Texas A&M University- Corpus Christi

PROFESSIONAL SERVICE

Mentor, Undergraduate Research Assistants Fall 2019—Present
Guided six RAs on projects as members of the "Motion Capture Crew" & "EEG Crew" within the Sensorimotor Neuroscience Lab, UC Merced
Cognitive & Information Sciences Graduate Group Administrative Fellowship Summer 2021
Updated and clarified policies and procedures on the CIS webpage, created FAQ and detailed guides for CIS graduate program milestones
Graduate Students Rights & Responsibilities Review Committee Spring 2021
Graduate Student Association, University of California, Merced
Professional Development Officer Fall 2019—Spring 2021
Cognitive and Information Sciences Graduate Student Group
University of California, Merced
Co-founder Cognitive & Information Sciences Graduate Student Group Fall 2019
University of California, Merced
Mentor, Women in Science and Technology Fall 2019—Spring 2020
WSTEM University of California, Merced
Delegate, Cognitive & Information Sciences Fall 2019—Spring 2020
Graduate Student Association, University of California, Merced
Co-organizer of Cognitive Science for the Common Good October 2019
Event Theme: Improving Workplace Climate & Gender Equity in Academia
Cognitive & Information Sciences, University of California, Merced

WORKSHOP & EVENT ORGANIZATION

Professional Development Event Series (Organizer) Fall 2019—Spring 2021

Industry(+) Job Market Panels (2)
Academic Job Market Panel
Giving Effective Presentations at Online Conferences (presenter)
Time Management Workshop (2)
Open Sciences & the Practicalities of Pre-registration
Graduate Student Experience & Navigating Support Structures
Grant Writing Workshop

Cognitive Science for the Common Good (Co-organizer) October 2019

Event Theme: Improving Workplace Climate & Gender Equity in Academia

Cognitive & Information Sciences Brownbag (Co-host) Spring 2019

Co-host of the CIS Graduate Student Speaker Series
Organized an Industry Jobs Panel w/ CIS Alumni
Organized an Academic Jobs Panel w/ Early Career CIS Professors

Interactive Alexander Technique Workshop March 2015

Organized a workshop enhancing awareness of the body in performance practice to prevent repetitive stress and posture related injuries and promote relaxed performance. *Guest speaker: Dr Teresa Ford, American Society for the Alexander Technique*

Interactive Dalcroze Eurythmics Workshop February 2014

Organized a workshop reinforcing rhythm, structure, and musical expression in the body through movement. *Guest speaker Dr David Frego, University of Texas at San Antonio*

Province Days: Three Day Educational Symposium February 2014

Co-organized musical performances, workshops on embodied musicianship and body awareness in musical performance, and presentations on arts management, arts communication, and financial management for performing artists. *Attended by South Texas Members of the Sigma Alpha Iota Music Fraternity and students from Texas A&M University-Corpus Christi, Del Mar College, Texas A&M University-Kingsville, and University of Texas-San Antonio*

SKILLS

Transcranial Magnetic Stimulation, EEG
Motion Capture (Vicon 3D MoCap system & integrated AMTI force plates; OpenPose 2D real-time multi-person keypoint detection)
R, Python, MATLAB

LANGUAGES

English (native)
French (speak, read, & write with proficiency)

ORGANIZATIONS

Society of Music Perception and Cognition
American Psychological Association
Cognitive Neuroscience Society

INSTRUMENTS

Primary: Oboe/English Horn
Secondary: Saxophone/Clarinet/Flute

Abstract

Rhythms at Small and Large Scales: the Neural Mechanisms of Rhythm Perception and the Recurrence Dynamics of Large Group Interaction

by

Shannon Proksch

Doctor of Philosophy in Cognitive and Information Sciences

University of California, Merced, 2022

Professor Ramesh Balasubramaniam, Ph.D., Chair

Recognizing that understanding the full scope of human cognition and behavior is intractable for any one discipline, the field of cognitive science has embraced the approaches and theoretical background of a diversity of fields, commonly listed as: philosophy, linguistics, anthropology, neuroscience, artificial intelligence, and psychology. This dissertation takes a similarly diverse approach to the study of music cognition and social interaction. In the first chapter, I review major theories in the neuroscience of musical rhythm perception. In the second, I conduct an experiment to determine how electrophysiological responses to musical rhythm are impacted by a brain stimulation method that down-regulates target regions of the brain. In the third chapter, I highlight how social interaction might involve coordination from low-level physiological signals up to high-level movement and acoustic signals that can be measured at the individual and group levels. In chapter four, I present a natural experiment applying nonlinear statistical analysis methods to acoustic data generated by a musical ensemble, identifying differing recurrence patterns dependent upon their mode of interaction. In chapter five, I extend these methods to a less scripted social interaction through analyzing the acoustic data generated by crowd sounds at a collegiate basketball game. This dissertation presents theoretical and experimental work spanning sensorimotor neuroscience, coordination dynamics, and complex systems. This work is intended to showcase a multiscale approach to the understanding of human cognition, applying multidisciplinary tools to questions within domains of music and social cognition.

Prologue

The brain and body are involved in a constant game of prediction. Our brain has certain expectations about the way the world will unfold around us. And our world is full of sensory signals soliciting certain patterns of expectancies and uncertainties in our brains. If met with unexpected input, our brains and bodies work together to form new expectations about the world. Music provides an exceptional playground of expectations and violations for our brains and bodies to exercise their predictive skills. Any musician could show you how playing with expectations and violations in melodic or rhythmic patterns will solicit different emotional responses or desires to move along to the beat. Some musical expectations might even be universal in different human cultures, and you don't have to be a musician to have developed these expectations already – the American singer Bobby McFerrin led a fantastic demonstration of one of these near universal musical expectations at the World Science Festival in 2009, in a now famous video about the pentatonic scale.[1]

If you watch that video, you will notice that McFerrin communicates a lot of information with his body, without saying much at all. He jumps in place to establish a rhythm. He also sings a note out loud, inviting the audience to sing that note to the beat of his jump. Before we even get to the second note, the audience has already had to perform a number of cognitive and perceptual exercises. Their brains had to build an expectation of the beat that they are seeing and hearing from the stage. Their visual and auditory systems had to communicate this sensory information from the stage to their motor systems, and use those sensory signals to entrain, or synchronize their voices to McFerrin's footfalls. The audience even accomplishes the challenging task of flexibly adjusting their singing to his changing tempos. (This tempo flexibility is something that many animals find particularly difficult). McFerrin then moves to the right and introduces a new note. The visual, auditory, and motor systems of the audience repeat this process of matching sensory expectations to the singing actions of their bodies. On stage, McFerrin jumps back and forth

---

[1] Bobby McFerrin at the 2009 World Science Festival: https://www.youtube.com/watch?v=ne6tB2KiZuk

between the two notes, and the audience is relying on sensorimotor couplings between their brain and body to sing along to the beat, when something exciting happens! McFerrin makes a jump to the right, and without prompting, the audience instinctively knows and sings the next note! Their brains predicted what note should happen on that jump, and since they did not *hear* that note from the stage, their bodies acted to reduce the uncertainty in the world—and the entire audience acted so as to make that note come about. They all *sang* the note they expected to hear next, and in doing so, their actions brought the world in line with their sensory expectations. And then, they keep doing it! The brains and bodies of the audience excel at this careful interplay of musical prediction and action as they continue singing a pattern of pitches and rhythms they have never practiced before, even as McFerrin improvises a tune on top of their voices, and even as he jumps to new 'notes' on the far left and right of the stage. If you were singing along (or imagining singing along) you may have had the same surprise as the audience in response to this amazing feat of prediction and action your brains and bodies are capable of!

This dissertation first presents research on how the brain processes rhythm – the regular organization of events in time – when we listen to music and rhythmic stimuli. In Chapter 1, I review research that reveals the integral role of the motor system in neural pathways that enable beat and rhythm perception. Much research in music neuroscience and cognition has asked, why should the motor cortex and other motor areas of the brain be important in perceiving a musical beat? When we watched Bobby McFerrin jumping on stage, he was using the movement of his body to show us the rhythm (and pitch), and we were using our motor system to interpret and translate his movements. Once idea is that maybe, we are using our motor cortex to simulate periodic motor planning activity – sort of like simulating actions that we would need to make in order to produce a certain rhythm – and we are using these simulations to help our auditory cortex predict when subsequent beats occur. Then the parietal cortex acts as the messenger, allowing the auditory and motor cortices to communicate with each other. Together, these brain regions make up the dorsal auditory pathway, which plays a pivotal role in our ability to perceive and produce rhythms. The Action Simulation for Auditory Prediction hypothesis posits that this pathway evolved in animals that have Vocal Learning abilities—such as humans and birds, but not monkeys—because vocal learning enhances tight connections between our auditory and motor cortices (Patel & Iversen, 2014). That is why humans and birds (but not monkeys) are pretty good at perceiving and producing rhythms. There's another hypothesis though, called the Gradual Audiomotor Evolution hypothesis, that posits that the evolution of beat perception and production in humans was a more gradual process, built on brain mechanisms that already exist in monkeys (Merchant & Honing, 2014). Monkeys can in fact perceive and produce some aspects of rhythms, like the length of a particular interval. But they can't process beat-based rhythms, like the recurring beat of a song. It turns out, in humans, there is a strong connection between the dorsal auditory pathway we discussed a moment ago, and the subcortical motor-cortico-basal ganglia-thalamo-cortical circuit. The strengthened connection between these two pathways may be the reason that humans are better perceivers and producers of rhythms than our primate relatives.

In Chapter 2, I experimentally investigate the role of two regions – the posterior parietal cortex (PPC) on the dorsal auditory stream, and the supplementary motor area which is not part of the dorsal auditory stream – in beat-based timing perception. Using a technique called Transcranial Magnetic Stimulation, I apply a magnetic field to the scalp to non-invasively stimulate electrical activity in the brain. The specific protocol we used, continuous Theta Burst Stimulation (cTBS) allowed us to downregulate neural activity in the stimulated brain region for a period of time. By down regulating a brain region, and comparing behavioral or neural activity before and after cTBS, we are able to make inferences about the causal role of that brain region in the activity of interest. We know that humans exhibit certain neural responses to rhythmic stimuli that are observable via Electroencephalography, or EEG. Specifically, if we play a predictable rhythmic pattern, but then deviate from that pattern for a single beat (a deviant stimuli), our brain recognizes this as an expectation violation which is reflected in an Event Related Potential (ERP). Participants in this experiment listened to rhythmic sounds with occasional deviant stimuli while we recorded brain activity with EEG. We used cTBS to down regulate either the PPC, the SMA, or we applied sham stimulation as a control. Comparing specific ERPs before and after brain stimulation allows us to evaluate the role of each brain area in rhythm perception. This chapter presents preliminary results on a subset of participants in a larger study.

Chapter 3 provides a general introduction to the second part of this dissertation, which investigates the emergent phenomenon of sensorimotor and neural entrainment extending over multiple bodies and brains. Recall the audience who participated in Bobby McFerrin's pentatonic scale demonstration. Specifically, recall how they coordinated with each other – singing the next note perfectly in unison, even though not one single audience member knew beforehand what note was coming next or when. In temporal rhythmic processing, the interaction of the body, brain, and environment result in an emergent phenomenon of sensorimotor and neural entrainment (J. M. Ross & Balasubramaniam, 2014). This emergent phenomenon is extended over multiple bodies and brains when the movement dynamics of one person becomes causally coupled with the movement dynamics of another person, forming an interpersonal synergy through their mutual interaction (M. A. Riley, Richardson, Shockley, & Ramenzoni, 2011). This can also occur when the acoustical dynamics of one person become causally coupled with the acoustical dynamics of other people. When we sing in unison, when we chant together at a sporting event, we form a group identity, merging together into a collective “we” that makes up a large interpersonal synergy (Cummins, 2020). The audience members attending McFerrin’s demonstration were all individually listening to the discussion on stage, maybe they were each fidgeting a bit in their seats or directing their attention to different speakers. But when McFerrin took the stage, directing everyone in the audience to sing along, they joined in a collective synergy feeding both off of McFerrin’s direction and on the sounds made by the rest of the audience. The audience members were coordinated across multiple signals – from neural, bodily, to social signals — contributing to their communication during that interaction (Pouw et al., 2021).

In Chapter 4, I demonstrate that we can take just this acoustic signal, the sound generated by an interacting group of people, and measure when they become coordinated in a single interpersonal synergy. Importantly, the individuals in this first group of people are members of a musical ensemble, musicians coming together to perform a shared musical work. By applying nonlinear statistical techniques to the audio signal generated by these musicians, we are able to differentiate when they are making music as a bunch of individuals sharing a stage, and when they have transitioned into a single musical ensemble. Just like the audience was aided by McFerrin’s direction when they coordinated into a single singing audience, our musicians were aided by a musical score, directing when they should begin coordinating. In Chapter 5, I demonstrate that we can evaluate differences in group acoustical interactions even when they are not being directed by something so explicit as a man on stage or a musical score. We apply the same analysis techniques to the sounds generated by fans at a collegiate basketball game engaging in collective behavior such as cheering on their team or making noise to distract the opposite team. The brains and bodies of the basketball crowd excel at a careful interplay of prediction and action as they join together in cheer and collective noise, falling in and out of interpersonal synergies with each other throughout the course of the game.

In the general discussion following these primary chapters, I contextualize the breadth of this work within an approach to cognitive science and the study of human experience that necessarily incorporates a multiscale approach in investigations of perception, cognition, and behavior in both music cognition and social interactions.

Chapter 1

Motor and Predictive Processes in Auditory Beat and Rhythm Perception

This chapter is published as a review in which two major hypotheses for the cortical and sub-cortical underpinnings of beat-based timing perception are compared. These hypotheses differ primarily in their views on how beat-based timing evolved and the reason we observe behavioral and perceptual differences between primates and humans in music and rhythmic timing tasks. This work is contextualized within recent action oriented approaches to predictive processing. The chapter ends with suggestions for future experiments to probe the causal role of regions in the proposed networks on the neural mechanisms underlying beat-based timing perception.

Published as:

Proksch, S., Comstock, D. C., Médé B., Pabst, A., Balasubramaniam, R. (2020) Motor and Predictive Processes in Auditory Beat and Rhythm Perception. *Frontiers in Human Neuroscience*

Abstract

In this article, we review recent advances in research on rhythm and musical beat perception, focusing on the role of predictive processes in auditory motor interactions. We suggest that experimental evidence of the motor system's role in beat perception, including in passive listening, may be explained by the generation and maintenance of internal predictive models, concordant with the Active Inference framework of sensory processing. We highlight two complementary hypotheses for the neural underpinnings of rhythm perception: The Action Simulation for Auditory Prediction hypothesis (Patel & Iversen, 2014) and the Gradual Audiomotor Evolution hypothesis (Merchant & Honing, 2014) and review recent experimental progress supporting each of these hypotheses. While initial formulations of ASAP and GAE explain different aspects of beat-based timing-–the involvement of motor structures in the absence of movement, and physical entrainment to an auditory beat respectively-–we suggest that work under both hypotheses provide converging evidence toward understanding the predictive role of the motor system in the perception of rhythm, and the specific neural mechanisms involved. We discuss future experimental work necessary to further evaluate the causal neural mechanisms underlying beat and rhythm perception.

Keywords: beat perception, motor system, motor planning, sensorimotor system, rhythm, timing

1.1 Introduction

The coupling of action and prediction in perception has been characterized by predictive models of perception (Rao & Ballard, 1999) (Rao and Ballard, 1999) including classical Predictive Coding (K. Friston, 2002, 2005) (PC), and the more recent Active Inference Framework (Active Inference – corollary to the Free Energy Principle, (K. J. Friston, Daunizeau, & Kiebel, 2009; K. Friston, 2010; Parr & Friston, 2019)). Under classical PC, the brain is thought to utilize an internal generative model and a process of probabilistic model updating to predict the causes of its sensory input. Each level of the neural hierarchy predicts the activity at the level below, with higher levels of the hierarchy providing empirical priors, or hypotheses that constrain the generation of new priors at the level below. At each level, the top-down predictive signal is compared to the bottom-up inputs from the lower level. When there is a mismatch between incoming, bottom-up sensory information and top-down predictions, a prediction error is propagated back to the level above where it is used to revise and improve the initial hypothesis. If the prediction error cannot be minimized at the level at which it is being processed, it is relayed up to the next level above. The higher in the hierarchy the prediction error is being relayed, the more substantial the revision in the hypothesis. Perceptual experience arises as prediction error is minimized and a ‘winning’ hypothesis is selected. Thus, the general idea of PC is perceptual inference.

However, this classical PC/Bayesian account of perception characterizes the brain as a passive, Helmholtzian, stimulus-response machine, responsive only to the generation of prediction errors between its top-down sensory predictions and the actual sensory input from the world (K. J. Friston & Stephan, 2007; Clark, 2013). Our brains are more aptly described as embodied and enactive, enabling us to move and interact with our environment to bring about the minimization of prediction errors through our own action (Thompson, 2010; Gallagher, Hutto, Slaby, & Cole, 2013; Bruineberg, Kiverstein, & Rietveld, 2018). This is the premise of Active Inference. As in PC, the brain uses an internal generative model to predict incoming sensory data. However, rather than relying on the passive accumulation of bottom-up sensory prediction errors that are minimized to create the content of perception, Active Inference formulations incorporate active engagement with the world to make the sensory inputs more predictable. Thus, in Active Inference, the prediction error minimization process which gives rise to perceptual experience is achieved through actions which conform sensory inputs to the brain's predictions (K. J. Friston et al., 2009; Hohwy, 2013; Parr & Friston, 2019).

Music perception and production are exemplar cognitive and behavioral phenomena to study these predictive processes and to evaluate the role of motor processing in sensory perception. Koelsch et al. (2019) expanded on the specific properties of music which make it an ideal paradigm for investigating predictive processes in the brain. Music, in any culture, is based on the generation of regularities, from the temporal regularities of rhythm to the predictable patterns and combinations of musical pitches. These regularities, or expectancies, generated by music have even been proposed as the properties which underlie emotional experience in music (Meyer, 1956; Huron, 2008; Juslin & Västfjäll, 2008). Cross-cultural perceptual priors may exist for some aspects of rhythm perception and production (Jacoby & McDermott, 2017), while other aspects are shaped by encluturation within a certain musical niche (Cameron, Bentley, & Grahn, 2015; van der Weij, Pearce, & Honing, 2017; Polak et al., 2018). In particular, the experience of musical groove, that property of 'wanting to move' to the music, is proposed to be related to the balance between prediction and prediction errors generated by rhythmic properties of the music (Janata, Tomic, & Haberman, 2012; Matthews, Witek, Heggli, Penhune, & Vuust, 2019; Matthews, Witek, Lund, Vuust, & Penhune, 2020). Active Inference formulations account for not only predictions related to expected stimulus input, but also predictions related to the expected accuracy—the precision, or uncertainty—of the original sensory prediction, in addition to counterfactual predictions related to how these prediction errors and their precision would change in response to active motor engagement with the sensory stimulus. Expected precision is modulated by sensory context and active engagement with the sensory signal. The generation of internal, predictive sensorimotor timing signals aligned to the musical beat may enhance the prediction and precision of temporal expectancies when perceiving syncopated musical rhythms, such as in musical groove (Koelsch, Vuust, & Friston, 2019). Whether or not we actually move our bodies to a musical rhythm, interactions between sensory and motor systems in our brain have been theorized to generate predictive timing signals that help us process musical rhythm (Merchant & Honing, 2014; Patel & Iversen, 2014; Vuust & Witek, 2014). These predictive timing signals are what allow for beat induction, or the active detection of the pulse in rhythmic time-varying stimuli such as music (Honing et al., 2014).

While predictive theories or perception are not new [indeed, they precede the age of Helmholtz, dating as far back as the 11th century works of Arab scholar al-Haytham et al. (ca. 1030;1989)], the purpose of this review is to contextualize recent advances in the role of the motor system in rhythm and musical beat perception under more recent advances within the Active Inference framework. We then directly compare two hypotheses for the neural underpinnings of rhythm perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis (Patel & Iversen, 2014) and the Gradual Audiomotor Evolution (GAE) hypothesis (Merchant & Honing, 2014). We suggest that the both hypotheses–taken together under the umbrella of Active Inference–provide converging evidence toward understanding the predictive role of the motor system within a distributed sensorimotor network underlying the perception of rhythm.

1.2 Action and Prediction in Rhythm Perception

The role of the motor system in rhythm perception is most obviously recognized by examining how it is we engage our body with music. In addition to beat induction in passive music listening, humans – and a limited group of birds and mammals (Kotz, Ravignani, & Fitch, 2018; Ravignani et al., 2019) – can move in time to a musical beat. This process of rhythmic entrainment is defined as the ability to flexibly perceive and synchronize to the beat of music or other complex auditory rhythms. It is argued that rhythmic entrainment abilities are determined by the ability to perceive a beat, the underlying pulse, within rhythmic stimuli. Beat perception in humans is inherently predictive, constructive, hierarchical, and modality biased. In addition, beat perception engages the motor system, even when no movement is present (Grahn & Brett, 2007; Chen, Penhune, & Zatorre, 2008a, 2008b; C. L. Gordon, Cobb, & Balasubramaniam, 2018)

In humans, behavioral evidence for prediction in beat perception comes from tapping experiments that reveal negative mean asynchronies, which are not observed in other primates. Asynchronies are observed when humans tap slightly earlier or later than the beat in a rhythmic stimulus, and negative mean asynchronies are a behavioral indicator that humans actively anticipate upcoming stimuli. Mean tapping asynchronies throughout a rhythmic stimulus are usually negative in the auditory domain, but much more variable in the visual domain (Pabst & Balasubramaniam, 2018). Humans also adjust future tapping response based on temporal mismatch between their movement and the current beat (Balasubramaniam, Wing, & Daffertshofer, 2004), and overtly tapping along to the beat aids in forming temporal predictions when compared to passively tracking a beat (Morillon & Baillet, 2017). In addition, when visual stimuli are presented motion (Hove and Keller, 2010) or a bouncing ball (Iversen, Patel, Nicodemus, & Emmorey, 2015), predictive entrainment as demonstrated by negative mean asynchrony becomes much more successful.

According to Active Inference, the brain minimizes prediction error either by updating predictions or by taking action in the world to bring actual proprioceptive input in line with top-down predictions regarding driving sensory stimuli. In musical beat perception, this means that we either take action and move to the beat, or we update our predictions by suppressing actual movement and instead establishing an internal model of the beat which corresponds to the proprioceptive input we would have received had we actually been moving to the beat. The ability to flexibly adapt motor behavior in response to a mismatch between a rhythmic auditory stimulus and current motor movement (Balasubramaniam et al., 2004) can be construed as one example of this more general active inference process. Enhanced rhythmic entrainment abilities for the visual domain when visual stimuli implies movement (Hove & Keller, 2010; Iversen et al., 2015), and the improvement of temporal predictions in conjunction with overt rhythmic movement (Morillon & Baillet, 2017) can also be explained by the increase of sensory information available in order to update and modulate descending predictions about the temporal regularities of the stimulus which guide motor movements.

But this Active Inference gloss on beat perception is - by itself – vague. Plausible neural architectures have been proposed to support the classical (Helmholtzian) PC/Bayesian processing of music in general (K. J. Friston & Friston, 2013). However, an empirically detailed account of the specific neural underpinnings of embodied Active Inference in human musical beat perception is necessary. The motor system has been proposed to play a key role in prediction and perception of sensory information (Schubotz, 2007), and is functionally organized to enable the driving (ascending) and modulatory (descending) message passing hypothesized within the Active Inference Framework (Adams, Shipp, & Friston, 2013). This differs slightly from traditional theories of motor control, where driving signals arise from descending, top-down motor commands. Under Active Inference, top-down predictive signals from the motor system serve to modulate proprioceptive predictions regarding driving, feed-forward sensory signals (Adams et al., 2013).

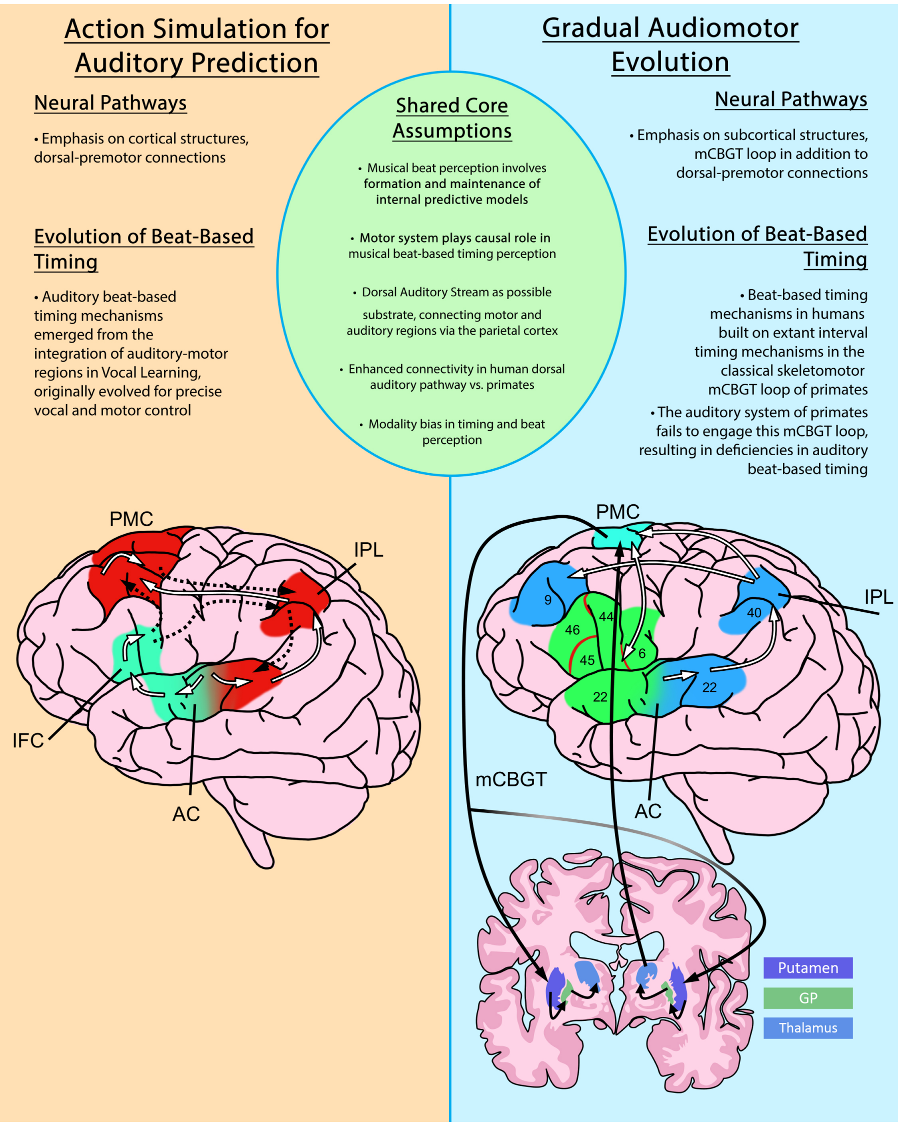
Concordantly, the motor system has been found to be consistently active when listening to music, even in the absence of specific motor movement. A recent meta-analysis of fMRI studies found clusters of activations in key regions of the motor system in passive music listening, including bilateral premotor cortex and right primary motor cortex (C. L. Gordon et al., 2018). Metrical musical stimuli have also elicited activation in the basal ganglia, supplementary motor area, and cerebellum (Grahn & Rowe, 2009). Indeed, the modality bias for human beat perception and rhythmic entrainment for auditory stimuli (Pabst & Balasubramaniam, 2018), and improvements of auditory beat processing when making overt action (Morillon & Baillet, 2017) can be explained by tight connections between auditory and motor regions of the brain. But the activation of motor structures of the brain, even in the absence of overt movement, indicates that the motor system plays a more fundamental role in the formation of abstract predictive models which support sensory perception (Schubotz, 2007; Adams et al., 2013; Patel & Iversen, 2014).

Strong explanations of rhythm perception must account not only for prediction in action, but also for the role of the motor activity observed in passive music listening. Below, we provide an overview on the motor system's role in rhythm perception, and review two complementary hypotheses which highlight the causal role of the motor system in beat-based timing perception.

1.3 Motor System in Rhythm Perception: Views from the Action Simulation for Auditory Prediction and the Gradual Audiomotor Evolution Hypothesis

Rhythm perception involves two types of timing perception,interval-based (absolute) timing and beat-based (relative) timing (Grube, Lee, Griffiths, Barker, & Woodruff, 2010; J. M. Ross, Iversen, & Balasubramaniam, 2016; J. M. Ross, Warlaumont, Abney, Rigoli, & Balasubramaniam, 2016; Iversen & Balasubramaniam, 2016). Interval-based timing refers to the ability to discriminate absolute differences in interval duration, whereas beat-based timing refers to the ability to measure the duration of time intervals relative to underlying temporal regularities such as beats (Teki, Grube, Kumar, & Griffiths, 2011). Beat-based timing perception is thought to be uniquely human (Merchant & Honing, 2014), and is believed to rely on the formation and maintenance of internal predictive models. According to the ASAP hypothesis (Patel & Iversen, 2014), these internal predictive models consist of periodic motor planning activity communicated via the dorsal auditory stream which allow for auditory prediction in beat-based musical timing perception. ASAP highlights the dorsal auditory stream due to its structural and functional relationship between auditory and motor planning regions, facilitating temporally-precise two-way signaling between these regions. This neural pathway involved in spatial processing of sounds (Rauschecker & Tian, 2000; Patel & Iversen, 2014) is more developed in humans than nonhuman primates, which is consistent with differences in beat-based timing behavioral ability (Honing, 2012; Patel & Iversen, 2014). In addition, Rauschecker (2018) postulates that the dorsal auditory stream may also be forming an "internal model of the outside world...[which] conver[ts] sensorimotor sequences into a unified experience"(p264–5). In the case of musical beat-based timing perception, the dorsal stream should form an internal model of the periodic musical beat.

Complementary to the ASAP hypothesis, the GAE hypothesis has been proposed to account for differences in beat-based temporal processing between primates and humans (see 1.1 for an overview comparison of ASAP and GAE). The GAE hypothesis (Merchant & Honing, 2014) also posits the dorsal auditory stream as a potential substrate for rhythm entrainment and perception. However, GAE claims that the evolution of rhythmic entrainment results more specifically from adaptations to the motor cortico-basal ganglia thalamo-cortical circuit (mCBGT). This specification arises from observations that the mCBGT is found to be active in sequential and temporal processing and movement in Macaques (Tanji, 2001; Merchant, Harrington, Meck, et al., 2013; Perez, Kass, & Merchant, 2013) and humans (Grafton, Hazeltine, & Ivry, 1995; Harrington, Zimbelman, Hinton, & Rao, 2010), including, for humans, the processing of musical rhythms (Grahn & Brett, 2007). Explicitly including the mCBGT loop in the evolution of rhythmic entrainment accounts for the fact that interval-timing ability appears preserved in macaques (Merchant et al., 2013) and is shared among primates, including humans. This indicates a shared neural circuitry for single interval-based timing, upon which GAE hypothesizes human beat-based timing mechanisms would have evolved to enable beat-based rhythmic entrainment. It is gradual changes to this foundational neural pathway, in addition to strengtheningconnections to auditory cortices via the dorsal auditory pathway, that have enabled the human mCBGT to develop beat-based timing mechanisms that can process the hierarchical properties of beat-based, rhythmic stimuli, such as music. Although focusing on slightly different neural pathways, both ASAP and GAE highlight the predictive role of the motor system in the perception of rhythm, and support growing consensus on the role of motor pathways in the formation of internal predictive models in perception more generally.



Action Simulation for Auditory Prediction

Neural Pathways

- Emphasis on cortical structures, dorsal-premotor connections

Evolution of Beat-Based Timing

- Auditory beat-based timing mechanisms emerged from the integration of auditory-motor regions in Vocal Learning, originally evolved for precise vocal and motor control

Shared Core Assumptions

- Musical beat perception involves formation and maintenance of internal predictive models
- Motor system plays causal role in musical beat-based timing perception
- Dorsal Auditory Stream as possible substrate, connecting motor and auditory regions via the parietal cortex
- Enhanced connectivity in human dorsal auditory pathway vs. primates
- Modality bias in timing and beat perception

Gradual Audiomotor Evolution

Neural Pathways

- Emphasis on subcortical structures, mCBGT loop in addition to dorsal-premotor connections

Evolution of Beat-Based Timing

- Beat-based timing mechanisms in humans built on extant interval timing mechanisms in the classical skeletomotor mCBGT loop of primates
- The auditory system of primates fails to engage this mCBGT loop, resulting in deficiencies in auditory beat-based timing

Figure 1.1: An overview comparison of the Action Simulation for Auditory Prediction Hypothesis (ASAP) and the Gradual Audiomotor Evolution Hypothesis (GAE). Shared core assumptions of both hypotheses are listed at center. Brief differing emphases on neural pathways and evolutionary commitments are listed in each panel. Diagrams depict the neural pathways proposed under each hypothesis. The ASAP diagram (left), shows ascending pathways from the auditory cortex (white lines) and descending pathways back to the auditory cortex (dashed lines) in the dorsal (red) and ventral (green) streams. The GAE diagram (right) shows the dorsal auditory pathway (white lines) and dorsal (blue) and ventral (green) streams, and the motor cortico-basal ganglia-thalamo-cortical (mCBGT) circuit (black lines). PMC, primary motor cortex; IPL, inferior parietal lobule; AC, auditory cortex; IFC, inferior frontal cortex; MPC, medial premotor cortex; GP, globus pallidus. Figures adapted from Merchant and Honing (2014) and Patel and Iversen (2014).

One important difference between the ASAP and GAE hypotheses is that ASAP purports to explain the presence of motor activity in beat perception even in the absence of overt movement, while GAE explains how evolution within motor pathways enables physical entrainment—synchronized movement—to a rhythmic stimulus. ASAP claims that beat perception in humans arose with the emergence of vocal learning abilities, which strengthened tight audio-motor connections in the dorsal auditory stream underlying rhythmic entrainment along the primate lineage.

In contrast, GAE favors a gradual strengthening of these connections over evolutionary time, building on specific interval timing mechanisms already extant in the mCGBT circuit of the primate brain. The result being the formation of an additional beat-based mechanism with enhanced connection of the mCGBT to the auditory cortex via that same dorsal auditory stream in the human brain (Merchant & Honing, 2014). Recent neurophysiological evidence highlights the interconnectedness of interval and beat-based timing mechanisms proposed by GAE, indicating that even in passive listening, monkeys are able to detect isochrony in rhythm, due in part to extant interval-based timing mechanisms of the monkey motor system, but that monkeys cannot detect the underlying beat in a rhythmic stimulus, which requires auditory-motor beat-based timing mechanisms present in humans (Honing, Bouwer, Prado, & Merchant, 2018).

1.4 Evidence for Prediction and Motor Activity in GAE and ASAP

In addition to fMRI observation of motor activation in music listening and rhythm processing, the predictive and causal roles of specific motor structures highlighted by the ASAP and GAE hypotheses have been experimentally tested via electroencephalography (EEG) and transcranial magnetic stimulation (TMS). Specific Event Related Potentials (ERPs) relating to prediction errors evoked by rhythmic deviations in musical stimuli include the mismatch negativity (MMN) and P3a (Honing et al., 2018; Koelsch et al., 2019). These auditory event related components indicate violation of temporal expectations in oddball paradigms, with early responses related to bottom-up sensory processing and later responses reflecting top-down cortical processes (Garrido et al., 2007) and (perhaps conscious) attention to deviant stimuli (Sussman, Winkler, & Schröger, 2003). EEG studies provide insight into the neural mechanisms of beat-perception while removing the limitations of behavioral response (Honing, 2012). The MMN and P3a components have been observed in response to rhythmic violations in adult humans, as well as infants and monkeys

(Ladinig, Honing, Haden, & Winkler, 2009; Winkler, Haden, Ladinig, Sziller, & Honing, 2009; Honing et al., 2018). However more recent research in monkeys comparing ERPs in passive listening to jittered and isochronous stimuli with occasional deviants have demonstrated that monkeys might be able to detect isochrony in rhythm – which could rely on extant interval timing mechanisms in the primate brain; but not the beat – which relies on more evolved beat-based timing mechanisms, while humans are able to detect both isochrony and the beat (Bouwer, Werner, Knetemann, & Honing, 2016; Honing et al., 2018). This collection of experiments supports the gradual evolution of beat-based timing mechanisms hypothesized by GAE.

Action Simulation for Auditory Prediction has been further supported by TMS research, demonstrating causal links between specific types of beat processing and regions of the dorsal auditory stream. A set of TMS experiments evaluated the role of the posterior parietal cortex (PPC), which is thought to serve as an interface for bidirectional communication between auditory and motor regions of the brain, and the dorsal pre-motor cortex (dPMC), which is also part of the dorsal auditory stream and is associated with movement planning and synchronization to auditory stimuli (Chen et al., 2008a; Giovannelli et al., 2014). By down-regulating neural activity in left PPC according to the Huang et al. (2005) protocol, Ross et al. (2018b) showed that left PPC may be involved in one aspect of beat-based timing–phase shift detection–but not tempo detection or discrete interval discrimination. Ross et al. (2018a) down-regulated activity in left dPMC, showing that left dPMC may be involved in tempo detection, but not phase shift detection or discrete interval discrimination. Additionally, measures of Motor Evoked Potentials (MEPs) in single pulse TMS over the motor cortex have indicated that musical groove modulates cortical excitability in the motor cortex. High levels of musical groove are characterized by syncopated rhythms, enhanced energy in the bass line, and the phenomenological property of 'wanting to move' with the music (Janata et al., 2012; Stupacher, Hove, Novembre, Schütz-Bosbach, & Keller, 2013; J. M. Ross, Warlaumont, et al., 2016). High-groove music has been shown to more strongly activate the motor system (resulting in higher MEPs) when compared with low-groove music (Stupacher et al., 2013). These results indicate the bidirectionality of auditory-motor interactions, as causally down-regulating activity in the motor cortex can impair auditory perception of aspects of musical rhythm, and varying degrees of rhythmic information in auditory stimuli (i.e., syncopation and bass frequencies in musical groove) can change aspects of motor cortical function.

1.5 Mechanisms for Timing and Rhythm Prediction

While there is growing consensus that the motor system is causally involved in timing and rhythm perception, and that the neural substrate includes cortical structures of the dorsal auditory stream and subcortical structures within the motor-cortico basal ganglia thalamo-cortical loop, the specific neural mechanisms which enable timing and rhythm perception within these substrates remains an open question. For some years, cognitive scientists have been looking for how internal timing can be instantiated by patterns of temporal stimuli via, e.g., clock-based or oscillatory mechanisms (Povel & Essens, 1985; Large & Jones, 1999). Given the amount of neuroscientific evidence pointing to a distributed timing network in the brain (Buonomano, 2014), mechanisms of entrainment to patterns of temporal stimuli have received significant attention. The striatal beat frequency model was suggested to support a clock-based mechanism based on banks of oscillators (Matell & Meck, 2000, 2004). In contrast, Large et al. (2015) describe an oscillatory model of pulse perception called Neural Resonance Theory (NRT), which provides a plausible mechanism of adaptive entrainment and beat-based timing without requiring an internal clock mechanism. According to NRT, rhythmic stimuli are encoded in sensory networks which interact with motor networks thus entraining them to the pulse frequency. Neural entrainment is induced to the pulse, even when the rhythmic stimulus itself lacks physical information at the location of the pulse—such as silences found ‘on the beat’ within syncopated rhythms—demonstrating the influence of top-down effects on pulse perception (Large, Herrera, & Velasco, 2015; Tal et al., 2017). The cerebellum has also been shown to play a prominent role in absolute timing (Nozaradan, Schwartze, Obermeier, & Kotz, 2017)–but not beat-based timing–with proposed mechanisms including an oscillatory pacemaker based on regular oscillations found within the inferior olive (Ashe & Bushara, 2014), and a state-spaced based mechanism, in which the timing of a stimulus can be inferred from the state of a relevant cortical network over time (Buonomano, 2014). In various cortical areas, ramping activity of neural firing rates has been proposed as a mechanism for interval-based timing—where interval duration is encoded in the modulation of neural spiking thresholds or by varying the slope of ramping activity preceding threshold (Durstewitz, 2003). However, in the Macaque brain, ramping activity has also been implicated for relative timing in coordination with multidimensional state space models as part of a multilayer timing system involving two neural populations (Merchant et al., 2014). These two neural populations are differentially associated with absolute and relative timing, and are observed in the medial motor cortex, consistent with the proposed role of the motor system under the GAE hypothesis (Crowe, Zarco, Bartolo, & Merchant, 2014; Merchant et al., 2014).

Continuous state-space models have also been proposed in Active Inference accounts for the generation of predictive models in action and sensory processing more generally, neurally mediated by the balance of pre- and post-synaptic activity (K. J. Friston, Parr, & de Vries, 2017) and neuronal firing rates in, e.g., medial or lateral intraparietal areas (de Lafuente, Jazayeri, & Shadlen, 2015). Striatal dopamine in particular has been proposed to code for both prediction error and certainty in response to sensory stimuli (Sarno, de Lafuente, Romo, & Parga, 2017) across a variety of timescales (Schultz, 2007). Dopaminergic activity also plays a role in rhythmic motor control (Koshimori & Thaut, 2018) and is responsive to rhythmic auditory stimulation (Koshimori et al., 2019), positioning dopamine as a crucial facilitator of the motor system's role in auditory-motor interactions underlying beat-based timing perception. The motor system's predictive role in music and rhythm perception is only one component of larger networks of sensorimotor processing, namely the dorsal auditory pathway and the mCBGT. Further experimental and computational work is necessary to determine whether and how the specific neural mechanisms of the human motor cortex processes timing information within the cortical and subcortical networks proposed by ASAP and GAE. To facilitate the generation of experimental and computational hypotheses, we have compiled an overview of recent experimental and theoretical research on the motor and distributed brain areas and mechanisms within the dorsal auditory pathway and the mCBGT–including the dopaminergic system–which are involved in the predictive processing of auditory-motor beat and rhythm perception in Table 1.1.[1]

---

[1] The motor system's predictive role in music and rhythm perception is only one component of larger networks of sensorimotor processing, namely the dorsal auditory pathway and the motor cortico-basal ganglia-thalamo-cortical circuit. This table provides an overview of the brain areas and mechanisms which make up these networks and are involved in the predictive processing of auditory beat and rhythm perception. Each brain area is introduced with one or more Theoretical or Review Papers contextualizing that brain area's proposed role, followed by a non-exhaustive list of supporting experimental work. This table is intended to serve as a tool for new or continuing researchers engaging in work on rhythm and musical beat perception.

Table 1.1: Overview of brain areas and mechanisms which make up cortical and subcortical networks involved in the predictive processing of auditory beat and rhythm perception

Brain Area	**Authors**	**Proposed Role of Each Brain Area**	**Experimental Task & Stimulus Type**	**Type of Data**
Cerebellum	Ivry & Schlerf, 2008	Dedicated timing mechanism; Coordination of movement, internal timing mechanisms involved with sub-second timing	Theoretical Paper/ Review	
	Bastian, 2006	Predictive models of movement	Theoretical Paper/ Review	
	Nozaradan et al., 2017	Tracking beats in rhythms with fast tempos; more prominent role in absolute timing vs relative timing	Passive Listening. Auditory rhythms designed to induce a beat - syncopated and unsyncopated.	EEG
	Gordon et al., 2018	Meta-analysis of fMRI studies of recruitment of motor system during music listening	Meta-analysis. Various listening tasks - Auditory rhythms or music.	fMRI
Basal Ganglia	Nozaradan et al., 2017	Tracking beats in complex rhythm sequences	Passive Listening Auditory rhythms designed to induce a beat - syncopated and unsyncopated.	EEG
	Merchant, Harrington, Meck, 2013	Interacts with the cortico-thalamic-striatal circuit in a context dependent manner	Theoretical Paper/ Review	
	Coull & Nobre, 2008	Perceptual temporal expectation; explicit timing	Theoretical Paper/ Review	
	Grahn, 2009	Internal beat generation; More prominent role in relative vs absolute timing	Discrimination task, same or different judgement of two auditory stimuli. Auditory rhythms - beat-based structure and non-beat-based structure; Accents- duration or volume accented (externally generated) or unaccented (internally generated) beats.	fMRI/ Behavioral
	Grahn et al., 2011	Internal representation of auditory rhythms that support cross-modal interactions in beat perception and generation	Discrimination task, rhythmic tempo change. Auditory tone metronome and visual flashing metronome. Two groups: one with auditory first visual second, and the other vice versa.	fMRI/ Behavioral

Brain Area	**Authors**	**Proposed Role of Each Brain Area**	**Experimental Task & Stimulus Type**	**Type of Data**
	Grahn & Rowe 2009	Internal beat generation: part of cortico-subcortical network involved in beat perception and generation	Indicate the strength of the percieved beat. Auditory rhythms of varying complexity and some with volume accents.	fMRI/ Behavioral
	Grahn & Rowe 2013	Putamen activity in beat prediction, but not beat finding	Attentive listening; occasionally indicate level of feeling of the beat. Auditory rhythms of varying intervals and rates, beat and non-beat (jittered) rhythms.	fMRI/ Behavioral
	Grahn & Brett, 2007	Higher activity for rhythms with integer ratio relationships between intervals and with regular perceptual accents	1st experiment (behavioral) reproduce auditory rhythms. 2nd experiment (fMRI) indicate if the rhythm played matched previous rhythms. Metered auditory rhythms of varying integer intervals and complexity.	fMRI/ Behavioral
	Teki et al., 2011	Striato-thalamo-coritcal network involved in beat-based timing, while an olivocerebellar network involved in duration-based timing	Judge duration matches in a set of tones. Auditory tones, either isochronous or jittered, arranged in in either rhythm-based or absolute duration-based sets.	fMRI/ Behavioral
	Araneda et al., 2017	Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream	Discrimination task, between beat and non-beat rhythms. Auditory, visual, and vibrotactile rhythms.	fMRI/ Behavioral
Primary and premotor cortices	Kilavik et al., 2014	Movement preparation, cue anticipation	Theoretical Paper/ Review	
	Schubotz, 2007	Predictive processing of external events, even in the absence of proprioceptive or interoceptive information	Theoretical Paper/ Review	
	Morillon & Baillet, 2017	Beta and delta oscillations directed to auditory cortex encode temporal predictions	Passive listening (listen condition); Active tapping with the beat (tracking condition). Auditory melody - different tones either on beat, anti-phase, or quasi-phase with the beat.	MEG/ Behavioral
	Gordon et al., 2018	Meta-analysis of fMRI studies of recruitment of motor system during music listening	Meta-analysis. Various listening tasks - Auditory rhythms or music.	fMRI

Brain Area	**Authors**	**Proposed Role of Each Brain Area**	**Experimental Task & Stimulus Type**	**Type of Data**
Premotor cortex	Grahn & Rowe 2009	Cortico-cortical coupling with SMA and Auditory cortex in duration beat perception; part of cortico-subcortical network involved in beat perception and generation	Indicate the strength of the percieved beat. Auditory rhythms of varying complexity and some with volume accents.	fMRI/ Behavioral
	Teki et al., 2011	Striato-thalamo-coritcal network involved in beat-based timing, while an olivocerebellar network involved in duration based timing	Judge duration matches in a set of tones. Auditory tones, either isochronous or jittered, arranged in in either rhythm-based or absolute duration-based sets.	fMRI/ Behavioral
	Chen et al., 2008	Motor regions recruited while listening to music rhythms	Experiment 1: Listen to rhythm passively then tap along with rhythm. Experiment 2: Listen to rhythm passively then tap along to rhythm without foreknowledge of being asked to tap with the rhythm Auditory tones in simply, complex, or ambiguous rhythms.	fMRI/ Behavioral
Supplementary motor area	Coull, Vidal, Burle, 2016	Perceptual and motor timing; Comparing the duration of perceptual events, error monitoring	Theoretical Paper/ Review	
	Ross, Iversen, Balasubramaniam, 2018	Not causally implicated in perceptual auditory interval timing	Discrimination task - same/different judgement of auditory intervals; Detection task - identification of tempo or phase shifted metronome click. Auditory intervals of pairs of tones; Metronome click track over musical stimuli.	Behavioral (pre/post TMS down-regulatory stimulation)
	Grahn & Brett, 2007	Higher activity for rhythms with integer ratio relationships between intervals and with regular perceptual accents; In musicians: higher activity for all rhythms when compared to rest	Experiment 1 (behavioral): reproduce auditory rhythms. Experiment 2 (fMRI): indicate if the rhythm played matched previous rhythms. Metered audtory rhythms of varying integer intervals and complexity.	fMRI/ Behavioral
	Grahn & McAuley 2009	Stronger activity in strong beat-perceivers vs weak beat-perceivers, no correlation with musicianship	Discrimination task, rhythmic tempo change. Auditory isochronous rhythms.	fMRI/ Behavioral

Brain Area	**Authors**	**Proposed Role of Each Brain Area**	**Experimental Task & Stimulus Type**	**Type of Data**
	Grahn & Rowe 2009	Coupling with STG in beat perception for musicians; part of cortico-subcortical network involved in beat perception and generation	Indicate the strength of the percieved beat. Auditory rhythms of varying complexity and some with volume accents.	fMRI/ Behavioral
	Teki et al., 2011	Striato-thalamo-coritcal network involved in beat-based time, while an olivocerebellar network involved in duration-based timing	Judge duration matches in a set of tones. Auditory tones, either isochronous or jittered, arranged in in either rhythm-based or absolute duration-based sets.	fMRI/ Behavioral
	Chen et al., 2008	Motor regions recruited while listening to music rhythms	Experiment 1: Listen to rhythm passively then tap along with rhythm. Experiment 2: Listen to rhythm passively then tap along to rhythm without foreknowledge of being asked to tap with the rhythm Auditory tones in simply, complex, or ambiguous rhythms.	fMRI/ Behavioral
	Araneda et al., 2017	Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream	Discrimination task, between beat and non-beat rhythms. Auditory, visual, and vibrotactile rhythms.	fMRI/ Behavioral
Medial premotor cortex	Merchant et al., 2014	Absolute and relative timing mechanisms within two separate neural populations	Theoretical Paper/ Review	
	Crowe et al., 2014	Absolute and relative timing mechanisms within two separate neural populations	Synchronization Continuation Task. Isochronous visual stimuli or auditory tones.	Behavioral; Extracellular activity of single neurons (in Macaca mulatta)
	Grahn & McAuley 2009	Stronger activity in strong beat-perceivers vs weak beat-perceivers, no correlation with musicianship	Discrimination task, rhythmic tempo change. Auditory isochronous rhythms.	fMRI/ Behavioral
Parietal Cortex	Patel & Iversen, 2014; Merchant & Honing, 2014; Rauschecker, 2011	Interface between motor and auditory cortices, sensorimotor integration	Theoretical Papers/ Reviews	
	Coull & Nobre, 2008	Perceptual temporal expectation; implicit timing	Theoretical Paper/ Review	

Brain Area	**Authors**	**Proposed Role of Each Brain Area**	**Experimental Task & Stimulus Type**	**Type of Data**
Auditory Cortex	Coull, Cotti, Vidal 2016	Temporal predictibility via fixed or dynamic predictions	Cued reaction time task. Visual cue that predicted target presentation time (temporal condition), or provided no information for target presentation (neutral condition) with variable intervals between cue and target.	fMRI/ Behavioral
	Ross, Iversen, Balasubramaniam, 2018	Causally implicated in perceptual beat-based timing	Discrimination task - same/different judgement of auditory intervals; Detection task - identification of tempo or phase shifted metronome click. Auditory intervals of pairs of tones; Metronome click track over musical stimuli.	Behavioral (pre/post TMS down-regulatory stimulation)
	Koelsch, Vuust, Friston, 2019	Event related potentials associated with predictive processes in music	Theoretical Paper/ Review	
	Fujioka et al., 2012	Beta-band activity predicts onset of beats in music	Passive listening, while watching silent videos. Auditory isochronous rhythms of several tempos and one irregular rhythm.	MEG
	Fujioka et al., 2015	Beta-band activity represents timing information being translated for auditory-motor coordination	Passive listening to metered rhythms, followed by attentive listening to un-metered rhythms that the participants were asked to imagine as metered. March and Waltz metered rhythms	MEG
	Auksztulewicz et al., 2010	Temporal prediction of rhythm and beats	Identify target chords. Auditory rhythmic or jittered sequences of distractor chords preceding target chords.	MEG/ EEG/ Behavioral
	Honing et al., 2018	Event related potentials to perceptual deviants in rhythmic stimuli	Passive listening. Auditory oddball paradigm with isochronous or jittered rhythms.	EEG (of Macaca mulatta)
	Bouwer et al., 2016	Event related potentials to perceptual deviants in rhythmic stimuli; ERPs modulated by attention in musicians	Passive or attentive listening. Auditory oddball paradigm with isochronous or jittered rhythms.	EEG

Brain Area	**Authors**	**Proposed Role of Each Brain Area**	**Experimental Task & Stimulus Type**	**Type of Data**
Dopaminergic System / Striatal Dopamine	Schultz, 2007	Multiple time courses of dopamine changes mediate multiple time courses of behavioral processes	Theoretical Paper/ Review	
	Friston et al., 2009	Reward learning, encoding of precision	Theoretical Paper/ Review	
	FitzGerald et al., 2015; Friston et al., 2012	Reward learning, encoding of precision	Theoretical Papers/ Computational Models	Simulated dopaminergic responses
	Sarno et al., 2017	Temporal expectation of perceptual cues; reward prediction error and (un)certainty	Detect weak vibrotactile stimuli. Variable interval durations between tactile start cue and vibrotactile stimuli.	Intracellular recording, monkey brain
	Koshimori et al., 2019	Rhythmic auditory stimulation (RAS) attenuates dopaminergic response	Synchronization task, RAS and no-RAS conditions; Various auditory rhythms, single auditory beats or metronome clickes over instrumental music.	Behavioral/ MRI/ PET
	Brodal et al., 2017	Rhythmic music reduces connectivity between basal ganglia and reward system	Passive listening. Electronic dance music in a continuous-stimulation design.	fMRI

1.6 Conclusions and Future Directions

In this paper, we reviewed recent advances in research on rhythm perception, focusing on the role of predictive processes in auditory motor interactions in beat-processing. We highlighted two complementary hypotheses for the neural underpinnings of rhythm perception: The ASAP hypothesis (Patel & Iversen, 2014) and the GAE hypothesis (Merchant & Honing, 2014) and reviewed recent experimental progress supporting each of these hypotheses. While initial formulations of ASAP and GAE explain different aspects of beat-based timing–the involvement of motor structures in the absence of movement, and physical entrainment to an auditory beat respectively–both theories have moved us closer to understanding the predictive role of the motor system in the perception of rhythm and the specific neural mechanisms involved. In fact, recent computational formulations of ASAP have further incorporated the subcortical structures proposed to be involved in the evolution of beat-based timing perception by GAE. Cannon and Patel (2019, preprint), have proposed the CBGT loop as responsible for the resetting of relative timing mechanisms via a hyper direct pathway from the SMA. In addition, they hypothesize a role for striatal dopamine in the maintenance of internal rhythmic timing models by tracking confidence (uncertainty) in the beat, consistent with Predictive Coding and Active Inference accounts of rhythm perception and perception more generally.

Future work in understanding the neural, cognitive, and behavioral dynamics of musical beat perception in humans should investigate not only the sensorimotor processes responsible for the perception of rhythm, but also the specific neural mechanisms by which top-down predictions serve to modulate driving proprioceptive sensations arising from concrete actions of the body or abstract activity of the motor systems. While EEG experiments (e.g., Ladinig et al., 2009; Winkler et al., 2009; Honing et al., 2018) point to the neural mechanisms of internal predictive models in beat-based timing perception, EEG alone cannot provide causal evidence forthe role of specific brain structures. Similarly, while TMS experiments (e.g., Stupacher et al., 2013; Ross et al., 2018a,b) have lended causal evidence for the role of specific structures in beat-based timing perception, the mentioned experiments do not provide direct evidence for the presence of internal predictive models of beat-based timing. If motor activity is causally involved in the formation of auditory predictions, then causal TMS manipulation to down-regulate activity in, e.g., parietal cortex or dPMC should result in the reduction of MMN and P3a event related responses to perceptual deviants in rhythmic stimuli, and this response might differ based on whether the stimuli contains timing deviants related to tempo or phase. Future research should include stimuli designed to elicit specific prediction errors with perceptual deviants, such as in oddball paradigms, while measuring event-related potentials associated with predictive processes in combined EEG and causal TMS experiments. Results from these experiments could extend and strengthen already emerging support for GAE and ASAP, as well as further contextualize the role of Active Inference in music and beat-based timing perception.

Author Contributions

SP and RB conceptualized the manuscript. BM, DC, and AP contributed to the writing and the exhaustive analysis of the literature. All authors contributed to the article and approved the submitted version.

Chapter 2

Causal TMS Investigation of Supplementary Motor Area and Posterior Parietal Cortex on ERP Responses to Auditory Rhythms

This chapter contains analysis of preliminary data from an unpublished and ongoing experiment probing the causal role of two cortical regions involved in beat perception. The experiment examines the effect of down-regulatory brain stimulation on well-known sensory evoked electrophysiological responses to deviants in rhythmic auditory stimuli. By comparing these evoked responses before and after brain stimulation, we can clarify whether or to what extent these cortical regions affect the low-level neural mechanisms involved in beat perception. Conclusions from these preliminary results indicate there may not be a large effect of stimulation, however final conclusions should not be drawn until complete data from the full experiment is assessed.

Authors:

Shannon Proksch and Ramesh Balasubramaniam *unpublished*

Abstract

Humans and rhesus monkeys (Macaca mulatta) share some aspects of rhythm cognition. Both monkeys and humans are sensitive to isochrony in a rhythmic stimulus, but only humans are sensitive to the metrical structure.

Humans remain sensitive to the isochrony and metrical structure of rhythmic stimuli even when the stimuli are unattended. Beat perception – the ability to perceive a regular pulse in (musical) rhythm – appears specific to humans. One reason for this may be stronger auditory-motor connections in the human brain compared to non-human primates. In this study, we evaluate the causal role of motor processes on auditory beat perception using electroencephalography (EEG) and transcranial magnetic stimulation (TMS). We presented human participants with isochronous (regular) rock beat stimuli with evenly spaced intervals and jittered rock beat stimuli with randomly spaced intervals, while preserving the order of individual sounds within each sequence. We compared ERP responses to oddball stimuli on the beat and off-beat in both conditions. We compared these same ERP responses before and after downregulatory TMS to either the supplementary motor area (SMA), posterior parietal cortex (PPC), or a sham stimulation condition.

2.1 Introduction

The human capacity for beat-based timing perception relies on auditory and motor predictive mechanisms (Balasubramaniam et al., 2021; Koelsch et al., 2019; Proksch, Comstock, Médé, Pabst, & Balasubramaniam, 2020; Cannon & Patel, 2021). The dorsal auditory stream, an auditory-motor network routed through parietal cortex, may be causally involved in making these sensory predictions (see Action Simulation for Auditory Prediction, ASAP (Patel & Iversen, 2014), as well as Gradual Audiomotor Evolution, GAE (Merchant & Honing, 2014)). Additionally, the supplementary motor area (SMA) is consistently active during musical and rhythmic tasks (C. L. Gordon et al., 2018) and is proposed to support timekeeping during beat anticipation (alongside subcortical structures(Merchant & Honing, 2014; Cannon & Patel, 2021)). Humans show stronger auditory-motor connectivity compared to non-human primates, which may explain better performance on behavioral beat-based timing tasks (Merchant & Honing, 2014).

Electroencephalography (EEG) experiments shed light on the neural mechanisms underlying rhythm perception that might not be apparent in behavioral tests. These studies have revealed event-related potentials (ERPs) relating to prediction errors as a result of rhythmic deviations in musical stimuli (Lumaca, Trusbak Haumann, Brattico, Grube, & Vuust, 2019; Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009) (see Koelsch et al., 2019 for a review). Two ERPs, the mismatched negativity (MMN) and P3a, are responses to temporal deviations in patterned stimuli that have been shown to differ in Rhesus macaques and humans when presented with on- or off-beat deviants in regular isochronous or jittered rhythmic stimuli (Bouwer et al., 2016; Honing et al., 2018). However, this EEG paradigm cannot isolate the causal role of specific neural structures involved. We seek to manipulate auditory-motor connectivity in human participants to reduce predictive timing and probe the causal role of PPC and SMA in beat-based timing.

Causal manipulation of neural activity, i.e. by downregulating neural firing using TMS, can be used to study the causal role of neural activity on subsequent perceptual or behavioral tasks (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005). TMS experiments have shown that downregulating neural activity in the SMA (involved in motor planning outside of the dorsal auditory stream) does not impact beat-based timing perception. However, downregulating activity in the PPC (which is part of the dorsal auditory pathway) does impact aspects of beat-based timing perception (J. M. Ross, Iversen, & Balasubramaniam, 2018; J. M. Ross, Proksch, Iversen, & Balasubramaniam, in review). The lack of effect from downregulating SMA on beat-based timing ability is surprising because SMA is consistently reported to be active in musical and rhythmic tasks, even in passive listening (C. L. Gordon et al., 2018). Furthermore, SMA (and its analogue, the medial premotor cortex in monkeys) is proposed to operate alongside subcortical structures to support timekeeping during beat anticipation (Cannon & Patel, 2021; Penhune & Zatorre, 2019; Gámez, Mendoza, Prado, Betancourt, & Merchant, 2019).

There is a known auditory bias for human rhythm perception in contrast to a visual bias for monkey rhythm perception. Behavioral data from monkeys engaging in simple synchronization continuation tasks demonstrates that monkeys have at least a precursor of human beat-perception, perhaps through extant interval timing mechanisms (Gámez et al., 2019; Penhune & Zatorre, 2019). Further, neural data from monkeys show MMN response to rhythmic grouping in isochronous stimuli and higher ERPs to deviants in isochronous compared to jittered stimuli, indicating a gradual evolution toward human auditory motor beat-based timing mechanisms as proposed by the Gradual Audiomotor Evolution (GAE) hypothesis (Merchant & Honing, 2014). Downregulating activity in the SMA or the PPC is expected to cause changes in the functioning of the neural mechanisms underlying human beat perception, reflected in ERP responses to perceptual deviants in rhythmic stimuli.

In contrast with GAE, the Action Simulation for Auditory Prediction (ASAP) hypothesis posits that human beat-based timing mechanisms arose from the coupling of auditory and motor processes as a result of vocal learning (Patel & Iversen, 2014). The cortical substrate supporting beat-based timing is the dorsal auditory pathway. This pathway implicates not the SMA, but the primary motor cortex (PMC). Motor planning activity in PMC provides predictive information for the motor cortex via the posterior parietal cortex. Thus, downregulating activity in the PPC – but not the SMA – is expected to cause changes in the neural mechanisms involved in beat-perception.

In this study, we further investigate the role of SMA and PPC in beat-based timing by combining the EEG paradigm and rhythmic stimuli from Bouwer et al., 2016 (Bouwer et al., 2016) and Honing et al 2018 (Honing et al., 2018) with the downregulatory TMS paradigm in Ross et al., 2018 (J. M. Ross et al., 2018). Observing changes in ERP response before and after downregulatory TMS stimulation will shed light on the causal roles of SMA and PPC. Incorporating EEG and a stimulus designed specifically to elicit neural prediction error responses will reveal changes in underlying neural mechanisms that may not have been observable in behavioral output of Ross et al., 2018. It is expected that the MMN and P3a ERPs will be attenuated after downregulatory TMS to parietal cortex or to SMA, such that prior to TMS ERP results should resemble those of human participants in Bouwer et al., 2016. Following downregulatory TMS, ERP results are expected to resemble those of the monkeys in Honing et al., 2018, i.e. with attenuated MMN and P3a response following TMS. These results will shed light on the neural mechanisms underlying human beat-based timing perception, and which differ between human and non-human primates.

2.2 Methods

Participants

Ten participants were recruited from undergraduate and graduate student population of the University of California, Merced. One participant's data was excluded due to extensive artifacts in their EEG signal, allowing us to analyze a total of nine participants for this study (average age = 22.7, female = 6) All participants were dominantly right-handed, with typical hearing, and screened for history of neurological disorders and other contraindications for TMS. The study was approved by the University of California, Merced, Institutional Review Board, and written informed consent was obtained from all participants.Total experiment time including consent and study preparation ranged from 2 to 2.5 hours per session. Participants were compensated $10 per each session.

Design

Participants completed three experimental sessions with a pre/post design in each session. Each experimental session consisted of:

1. Pre-stimulation (baseline) EEG recording during passive listening to rhythmic sequences

2. Administration of cTBS to target location
3. Post-stimulation EEG recording during during passive listening to the same rhythmic sequences

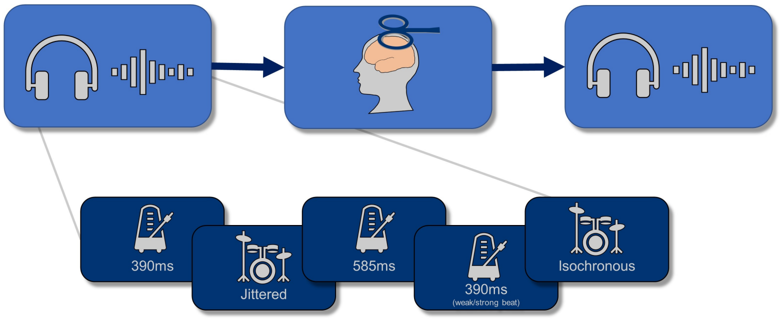


Figure 2.1: Study Design. Each participant listened to five acoustic stimuli, received TMS, then heard the same acoustic stimuli in the same order. Stimuli were randomized across sessions.

As a within-subjects design, each participant received TMS to each target location (sham, lPPC, and lSMA) on three separate days with EEG recording sessions pre- and post-TMS each day. Each EEG recording session consisted of five passive listening blocks: two ten minute blocks of jittered or isochronous rock beats, and three 3 to 7 minute blocks of metronome tones. All five blocks were randomized and counterbalanced across participants, with each participant completing all five blocks before and after TMS each session. Within one session the order of the blocks stayed the same pre and post-TMS, but the order of blocks across sessions were randomized each day. Only the Jittered and Isochronous rock beat conditions are evaluated in this study.

Stimuli

Rhythmic rock beat stimuli were a subset of stimuli from Bouwer et al, 2016 (Bouwer et al., 2016) (also used in Honing et al 2018), and the following description of the stimuli was adapted from those two papers. The rhythmic sequences were composed of two sounds: a simultaneously sounding bass-drum and high-hat sound (i.e. accented; 110ms duration), and a hi-hat sound (i.e. unaccented; 70ms duration). The accented sound was 16.6db louder than the unaccented sound, and together these acoustic features create a salient pattern of accented and unaccented beats. Deviant sounds were created by attenuating the accented bass-drum+hi-hat sound by 25dB (i.e. attenuated;

110ms). These sounds were combined in four patterns with two-sounds each, which were concatenated into longer 9minute 45 second rhythmic sequences. Standard pattern S1 contained an accented followed by an accented sound and occurred for 60% of the rhythmic sequence, while standard pattern S2 contained an accented sound followed by a second accented sound which occurred for 30% of the sequence. The remaining 10% of the rhythmic sequence contained 'oddball' stimuli: an attenuated sound followed by an accented sound (deviant pattern D1; 5% of the sequence) or an accented sound followed by an attenuated sound (deviant pattern D2; 5% of the sequence). These patterns were combined into regular and jittered conditions. Regular conditions contained consistent 225ms intervals, inducing a beat with a 450ms interval. Jittered conditions contained intervals randomly distributed between 150 and 300ms with an average of 225ms (uniform distribution), to make beat perception impossible. However, the inter onset interval immediately before and after a deviant tone was kept constant at 225ms to keep the acoustic and temporal context surrounding a deviant identical between both conditions for subsequent ERP analysis.

Concatenation of patterns in each rhythmic sequence was semi-randomized according to four constraints: 1) Pattern S2 was never presented more than once in a row; 2) a maximum of four consecutive S1 patterns was allowed; 3) a deviant on beat (D1) always followed an unaccented offbeat (S2); 4) at least five standard patterns were presented between two deviant patterns. Sounds are referred to as on and off beat sounds in both the jittered (non-isochronous) as well as the regular (isochronous) condition, although they are perceivable as on/off beats only in the isochronous condition. See Honing et al 2018 or Bouwer et al 2016 for a visual description of the stimuli. A recording of the stimuli used can be found in Supplementary Materials.

To fit within the effective window of cTBS, a single 9m 45s regular sequence and a single 9m 45s jittered sequence were presented to each participant before and after cTBS, in addition to the three metronome stimuli not presented in this paper. Pilot data consisting of three participants suggested this was sufficient to elicit ERPs. Stimuli were presented using Paradigm experimental stimulus presentation software (Perception Research Systems, 2007). Auditory stimuli were sent over headphones, while participants fixated on a single image presented on a 60Hz monitor a comfortable distance away. Overall, the study design included 19m 30s of rock beat stimuli (jittered and regular blocks) plus 13m 30s over three blocks of metronome stimuli. Participants were allowed short breaks between blocks, resulting in an average of 37 minutes for each EEG recording session pre and post TMS. Total experiment time including consent and study preparation ranged from 2 to 2.5 hours each day.

TMS Procedure

Each participant completed passive listening tasks while recording with EEG (as described in section XX) before and after TMS. Downregulatory TMS was applied using continuous-theta burst stimulation (cTBS) following the protocol from Huang et al., 2005 (Huang et al., 2005) with the Magstim Rapid2 system. Specifically, cTBS was applied in bursts of three pulses at 50Hz, repeated at 200ms intervals, for a total of 600 pulses over 40 seconds at 80% of each participant's active motor threshold (AMT). If a participant's 80% of AMT was a greater intensity than can be safely administered by our machine, then we administered cTBS at the greatest intensity that was safe (45% of maximum stimulator output).

AMT was determined as the lowest intensity observed to produce a visible twitch in 5 out of 10 trials in the flexed first dorsal interosseous (FDI) muscle with single pulse to a left primary motor cortex hotspot. Visible twitch was verified by motor-evoked potentials (MEPs) of at least 50 microvolts measured from the right-FDI muscle when administering single pulse TMS over the motor hotspot. The best location for the motor hotspot was determined by comparing MEP size and consistency at rest. MEPs were measured using surface electrode myography (EMG) with Ag/AgCl sintered electrodes over the belly of the right FDI muscle with a ground electrode placed over bone near the right elbow. Single pulse TMS to primary motor cortex was conducted using a figure-of-eight coil (Magstim, D702 70mm coil, Carmarthenshire, United Kingdom) held tangential to the scalp surface at an angle of 45 degrees from the anterior-posterior midline.

Neuronavigation to lPPC and lSMA was achieved using Magstim Visor2 3-D motion capture-guided neuronavigation. Each participant's brain model was scaled to the Talairach brain using head size and shape. Coordinates for lPPC and lSMA stimulation cites were determined from previous literature. The lPPC target was Talairach -40, -50, 51 following Krause et al 2012 and Ross et al 2018, and the lSMA target was Talairach -6, -12, 54 following Chauvigné et al 2014 and Ross et al 2018. Coil orientation for cTBS was 45 degrees from the anterior-posterior midline with the coil facing anterior and held tangential to the scalp (Janssen, Oostendorm and Stegeman, 2015. Sham cTBS was administered over left M1 with the coil facing away from the participant's head. Each experimental session was separated by at least two days, with no more than two sessions per week following UC Merced IRB protocol for TMS experiments.

EEG Data Acquisition and Processing

Electroencephalography was collected using an ANT-Neuro 32 channel amplifier with the ANT-Neuro 32 electrode Waveguard cap. Electrodes were positioned according to the 10-20 international system, and EEG data were recorded with a sampling rate of 1024Hz. Data were pre-processed using the EEGLab toolbox (version 2022.0; (Delorme & Makeig, 2004)) and ERPLab toolbox (version 9.00; (Lopez-Calderon & Luck, 2014)) with MATLAB 2018b. Channel locations were added using the standard location montage for the Waveguard cap.

Data were downsampled to 512Hz, high-pass filtered with a 0.1Hz frequency cutoff IIR Butterworth filter with a filter order of 4 and 24-80dB roll-off, and DC offset removed. ICA was run to remove eyeblinks and saccades for each participant before data was re-referenced using the average reference. Frontal electrodes (FP1, FPz, and FP2) for three participants were removed prior to ICA and interpolated following ICA component rejection. Data was epoched to a range of -150 to 500ms for all channels before and after standard and deviant stimuli, with pre baseline correction. Epochs with deflections exceeding +/- 150 microvolts were rejected, and data were filtered with a low-pass IIR Butterworth filter at 30Hz. Difference waves were obtained by subtracting ERP responses to the standard sounds from ERP responses to the deviant sounds at the same position (beat or off beat). We averaged over all participants to obtain grand average ERPs and difference waves.

EEG Data Analysis

Two clusters were identified in the grand average ERPs: a negative right fronto-central cluster and a later positive fronto-central cluster. We defined region of interest across these clusters as shown in Table 2.1. Difference waves measured across five electrodes in the right cluster exhibited an early negative peak between 100 and 200 ms after the onset of the deviant consistent with the latency of the MMN. Difference waves measured across six electrodes in the central cluster exhibited positive peaks observed between 200 and 350 ms after onset of the deviant consistent with a the latency of the P3a and P3b components. Following Bouwer et al., 2016 we defined ERP components as a 60 ms window surrounding peak amplitudes corresponding the MMN and P3a.

Right Cluster	Central Cluster
–	F3
Fz	Fz
F2	F2
–	FC1
FC2	FC2
Cz	Cz
C4	–

Table 2.1: Electrodes analyzed for the MMN component (Right Cluster) and the P3a component (Central Cluster).

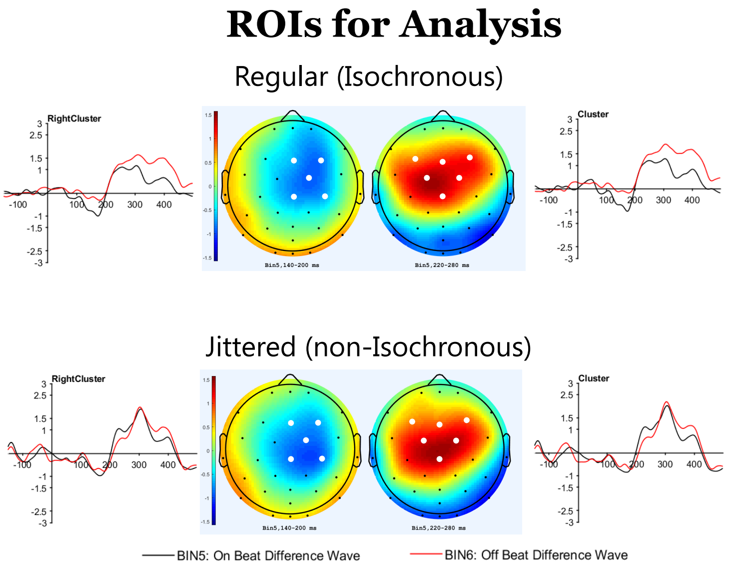


Figure 2.2: Topoplots show average activation in response to an "on beat" deviant within a 140 to 200 ms window corresponding to the MMN (Right Cluster), and a 220 to 280 ms window corresponding to the P3a (Central Cluster). Line plots show difference waves for both "on" and "off" beat deviants in each ROI cluster.

The observed peak negative latency fell at about 174ms, allowing a window between 144 to 204 ms. To ensure we did not overlap with the following positive component, we offset this window by -5 ms. Thus, we defined the MMN as the average amplitude in a sixty second window between 140 and 200ms. There were two peaks in the P300 range at 250 and 305 ms. We chose a 60ms window surrounding the first peak to focus on the P3a component, and defined the P3a as the average amplitude between 220 and 280ms.

Statistical Analysis

To test the effect of regularity and beat position in the baseline data, mean ERP amplitudes extracted from difference waves were compared using a linear mixed effects model with fixed effects of regularity and position, and a random effect of subject. This random effect accounts for each participant having a different intercept. A separate model was run for each ERP component (MMN and P3a).

- Amplitude ∼ Regularity (Reg/Jit) * Position(on/off beat) + (1| subj)

To test the effect of cTBS stimulation, mean ERP amplitudes extracted from difference waves were compared using a separate linear mixed effects model for each stimulation condition. The model for each stimulation condition contained fixed effects of stimulation, regularity, and position, with a random effect of stimulation by subject. This random effect structure allows each participant to have both varying intercepts and varying slopes. A separate model was run for each stimulation site (Sham, PPC, SMA) and for each ERP component (MMN and P3a).

- Amplitude ∼ Stimulation(Pre/Post) * Regularity (Reg/Jit) * Position(on/off beat) + (stim | subj)

Finally, to compare across cTBS stimulation sites, a pre minus post-cTBS difference score of mean ERP amplitude was calculated for each cTBS condition and compared using a linear mixed effects model with fixed effects of stimulation site, regularity, and position with a random effect by subject. A separate model was run for each ERP component (MMN and P3a).

- Amplitude difference score ∼ Stimulation Site(Sham/PPC/SMA) * Regularity (Reg/Jit) * Position(on/off beat) + (1| subj)

2.3 Results

Baseline ERP Reults across all sessions

Results from this preliminary data over all baseline conditions partially replicate Bouwer et al 2016. The mismatched negativity (MMN) ERP response to deviant stimuli was highest (most negative) for on beat sound events in the regular (isochronous) rhythmic sequence compared to off beat events in the regular sequence. MMN for on beat and off beat events in the jittered rhythmic sequence did not differ. The later P3a component showed higher mean amplitude for off beat deviants than on beat deviants in the isochronous condition. As expected, the P3a was lower for off beat deviants than on beat deviants in the jittered condition. Difference waves and topoplots showing mean activation in each ROI can be seen in Figure 2.3 and Figure 2.4. Mean ERP responses are showin in Figures 2.7 and 2.8. Linear mixed effects model results evaluating the effect of regularity and beat position for both MMN and P3a amplitudes for all baseline conditions can be seen in Table 2.2.

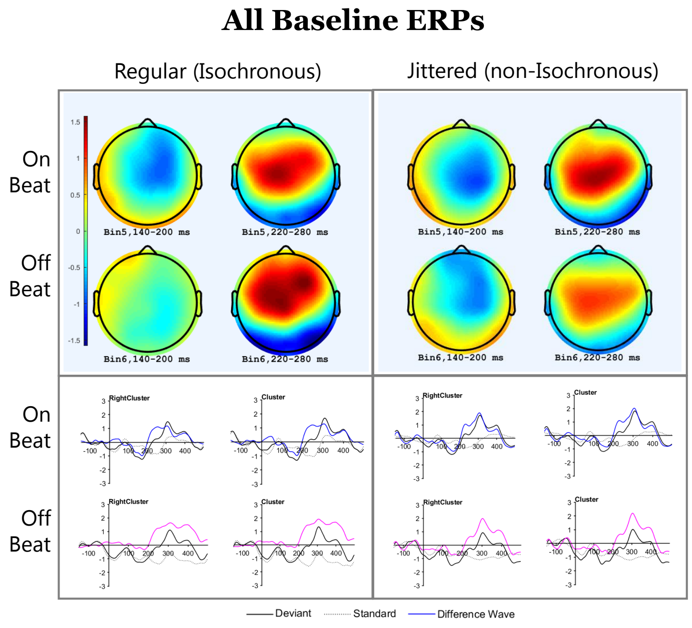

All Baseline ERPs

Figure 2.3: Topoplots show average activation in response to “on” and “off” beat deviants. Line plots show standards, deviants, and difference waves measured from the corresponding ROI for both “on” and “off” beat deviants. Data correspond to the baseline EEG recording (pre-TMS) over all three experiment sessions.

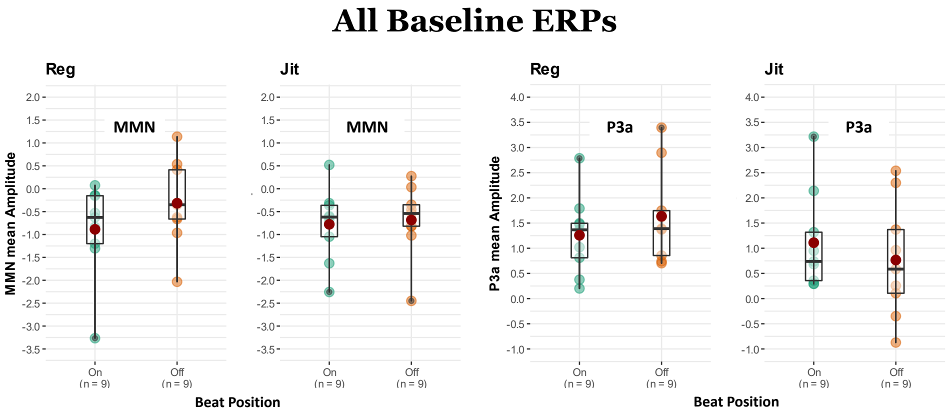


Figure 2.4: Mean ERP amplitude in response to "on" and "off" beat deviants. Data correspond to the baseline EEG recording (pre-TMS) over all three experiment sessions.

Table 2.2: Linear Mixed Effects Model Results – Baseline ERPs

MMN Amplitude ∼ Stimulation (Pre/Post) *
Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-0.89	-1.46 – -0.32	**0.003**
Jitter	0.11	-0.24 – 0.46	0.531
OffBeat	0.57	0.22 – 0.92	**0.003**
Jitter:OffBeat	-0.48	-0.98 – 0.02	0.062
σ^2	0.13		
Marginal R^2	0.063	Conditional R^2	0.819

P3a Amplitude ∼ Stimulation (Pre/Post) *
Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	1.26	0.63 – 1.88	**<0.001**
Jitter	-0.15	-0.64 – 0.35	0.547
OffBeat	0.38	-0.12 – 0.87	0.131
Jitter:OffBeat	-0.72	-1.42 – -0.02	**0.044**
σ^2	0.26		
Marginal R^2	0.106	Conditional R^2	0.719

ERP Results Pre and Post TMS

We analyzed mean ERP amplitude pre to post-cTBS within each stimulation site. Difference waves and topoplots showing mean activation in each ROI can be seen in Figures 2.7 and 2.8 respectively. The preliminary results for the linear regression model including the effects of pre to post cTBS stimulation are shown in Tables 2.3, 2.4, and 2.5 respectively.

We expected to find no significant differences pre and post sham stimulation, however the interaction effect of beat(off beat) on stimulation(post) * regularity(jitter) was statistically significant and positive (beta = 1.62, 95 % CI = -0.09 - 3.13, p = 0.038), representing an increase in P3a mean amplitudes for that interaction compared to the intercept of stimulation(pre)*regularity(regular)*beat(on beat). All other differences were non-significant for both MMN and P3a mean amplitudes for sham stimulation.

We hypothesized attenuated MMN and P3a mean amplitudes pre to post either PPC and SMA stimulation. For PPC stimulation, the effect of beat(off beat) on P3a mean amplitudes was statistically significant and positive (beta = 0.67 , 95% CI = 0.03 – .321, p= 0.040), representing a decrease in P3a for off beat deviants. The interaction effect of beat(off beat) on regularity(jitter) was statistically significant and positive (beta = 0.71, 95% CI = -0.58 – 1.99, p= 0.278), representing an increase in the P3a for off beat deviants in the jittered condition. All other differences were non-significant for both MMN and P3a mean amplitudes for PPC stimulation.

For SMA stimulation, the effect of beat(off beat) on MMN mean amplitudes was statistically significant and positive (beta = 0.65, 95% CI = 0.11 – 1.20, p = 0.019), representing a decrease (less negative) in MMN for off beat deviants. All other differences were non-significant for both MMN and P3a mean amplitudes for PPC stimulation.

Table 2.3: Linear Mixed Effects Model Results – Pre and Post Sham cTBS Stimulation

MMN Amplitude ∼ Stimulation (Pre/Post) * Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-1.07	-1.73 – -0.41	**0.002**
Post	0.47	-0.24 – 1.18	0.192
Jitter	0.50	-0.11 – 1.11	0.105
OffBeat	0.54	-0.06 – 1.15	0.078
Post:Jitter	-0.52	-1.38 – 0.33	0.022
Post:OffBeat	-0.61	-1.47 – 0.24	0.157
Jitter:OffBeat	-0.65	-1.51 – 0.21	0.136
Post:Jitter:OffBeat	1.11	-0.10 – 2.33	0.071
σ^2	0.41		
Marginal R^2	0.048	Conditional R^2	0.581

P3a Amplitude ∼ Stimulation (Pre/Post) * Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-1.05	0.34 – 1.76	**0.005**
Post	0.58	-0.22– 1.38	0.154
Jitter	0.06	-0.70 – 0.82	0.875
OffBeat	-0.01	-0.77 – 0.75	0.983
Post:Jitter	-0.78	-1.85 – 0.30	0.154
Post:OffBeat	-0.77	-1.85 – 0.30	0.156
Jitter:OffBeat	-0.39	-1.46 – 0.69	0.476
Post:Jitter:OffBeat	1.62	-0.09 – 3.13	**0.038**
σ^2	0.65		
Marginal R^2	0.062	Conditional R^2	0.470

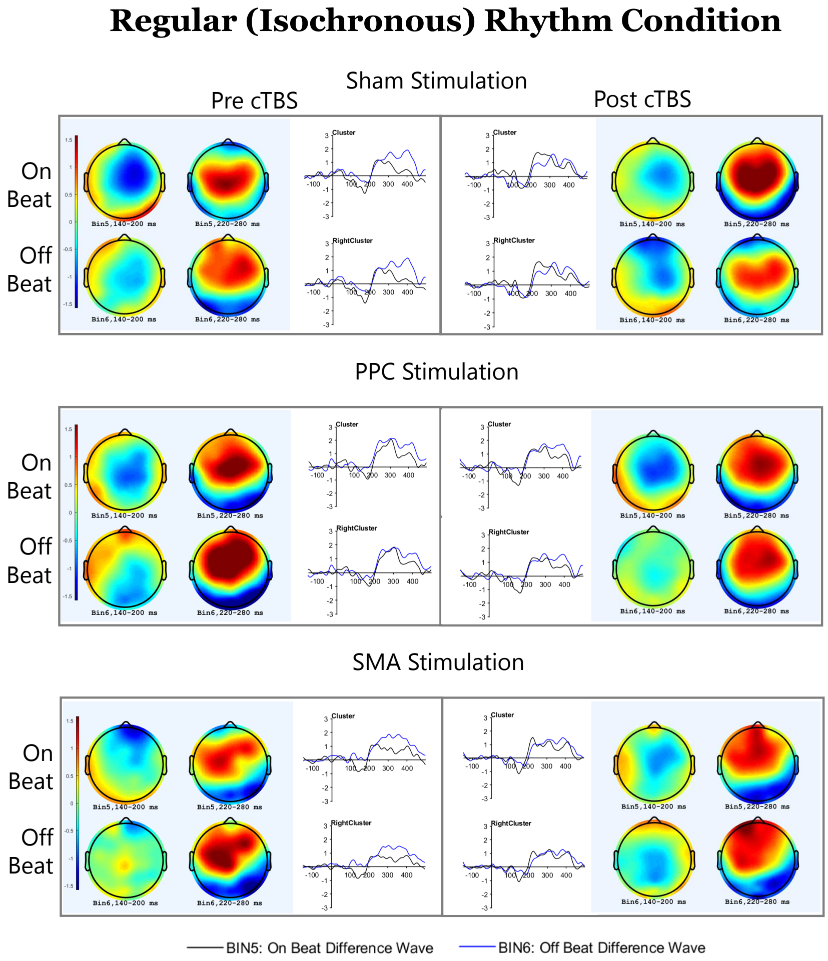


Figure 2.5: Topoplots show average activation in response to “on” and “off” beat deviants. Line plots show standards, deviants, and difference waves measured from the corresponding ROI for both “on” and “off” beat deviants. Data correspond to the the regular (isochronous) rhythm condition, pre and post cTBS at each stimulation site.

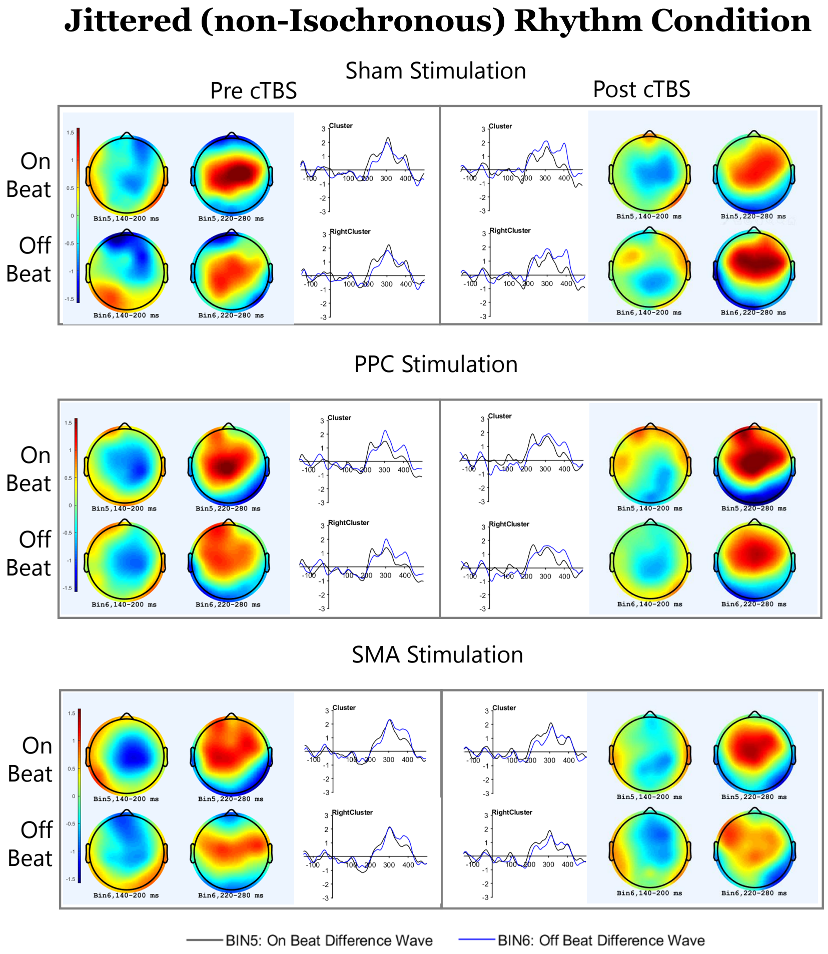


Figure 2.6: Topoplots show average activation in response to “on” and “off” beat deviants. Line plots show standards, deviants, and difference waves measured from the corresponding ROI for both “on” and “off” beat deviants. Data correspond to the the jittered (non-isochronous) rhythm condition, pre and post cTBS at each stimulation site.

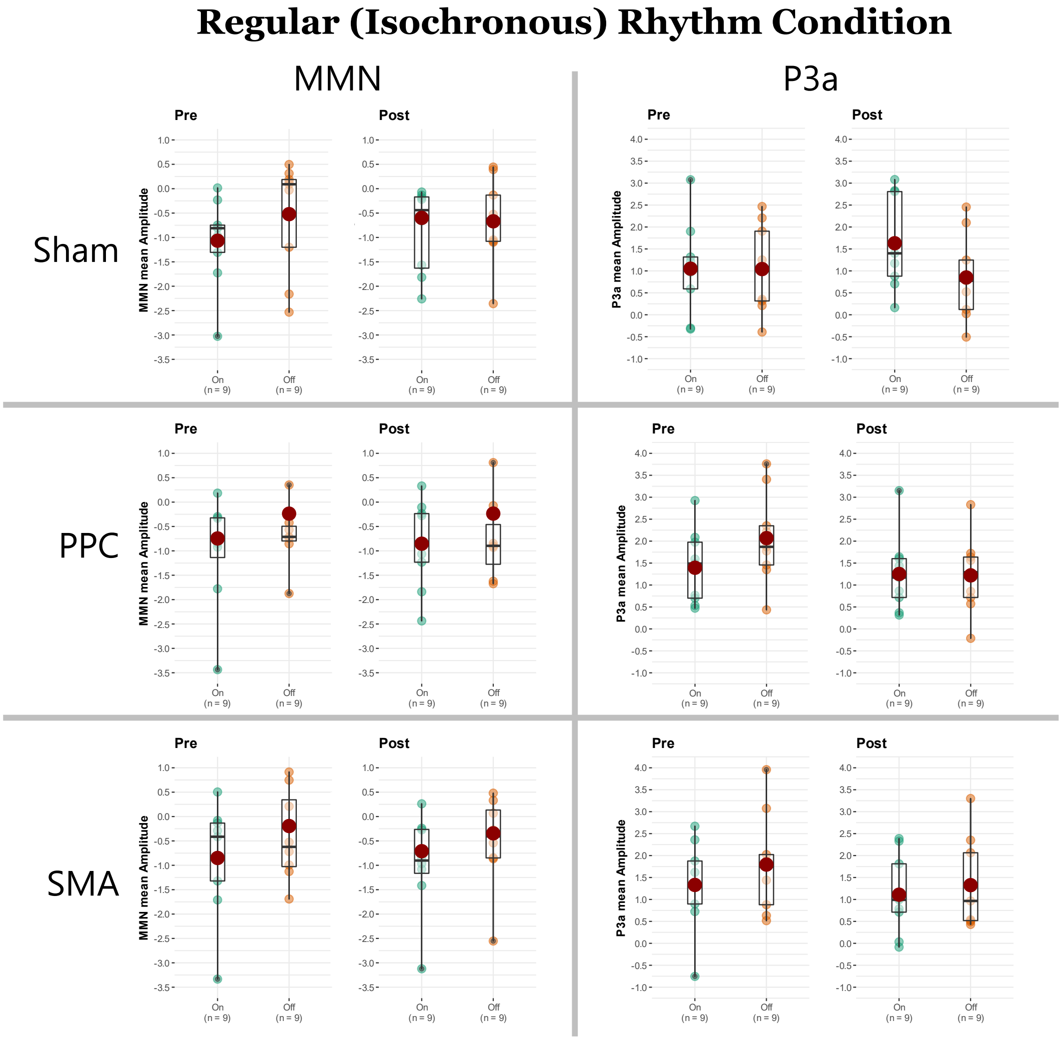


Figure 2.7: Mean ERP amplitude in response to “on” and “off” beat deviants. Data correspond to the the regular (isochronous) rhythm condition, pre and post cTBS at each stimulation site.

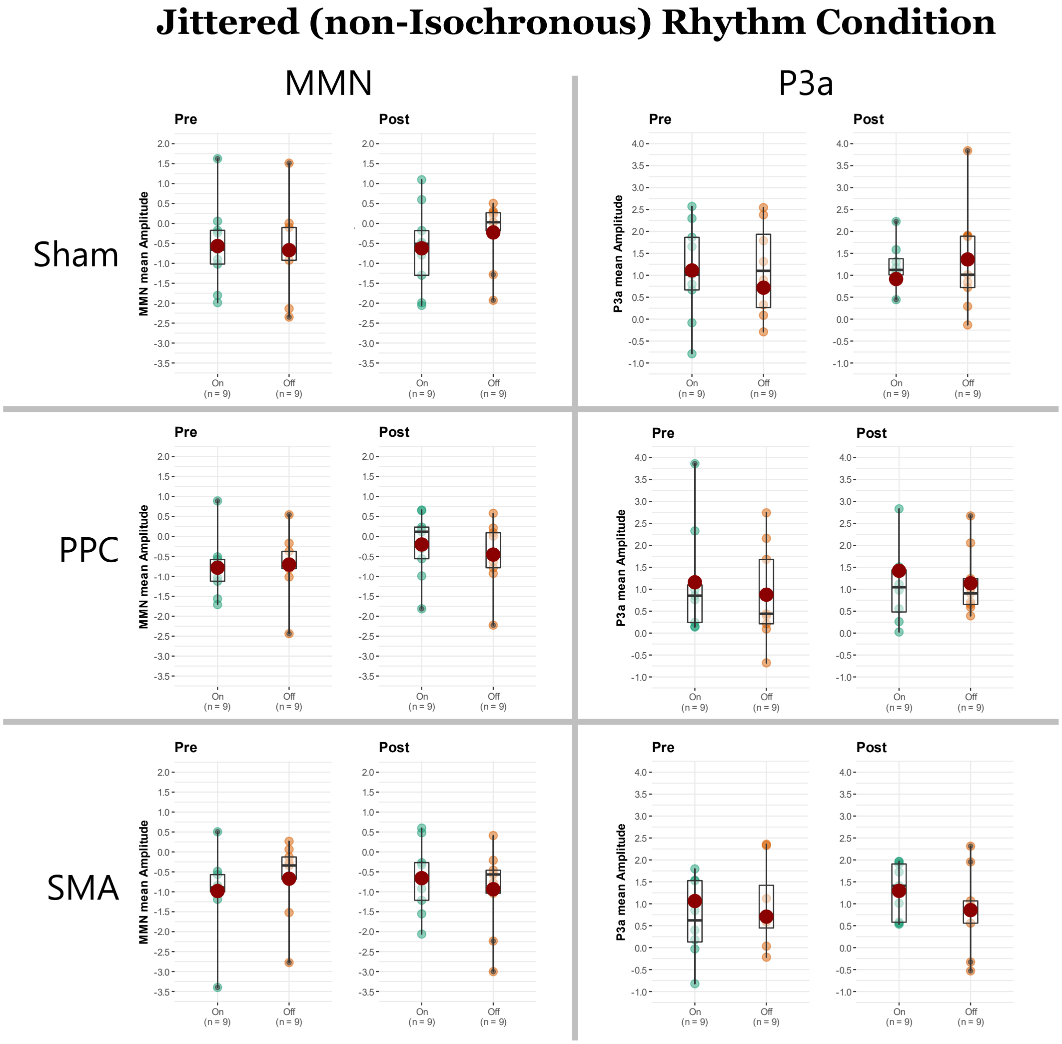


Figure 2.8: Mean ERP amplitude in response to "on" and "off" beat deviants. Data correspond to the the jittered (non-isochronous) rhythm condition, pre and post cTBS at each stimulation site.

Table 2.4: Linear Mixed Effects Model Results – Pre and Post PPC cTBS Stimulation

MMN Amplitude ∼ Stimulation (Pre/Post)
* Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-0.75	-1.39 – -0.10	**0.024**
Post	-0.11	-0.67– 0.45	0.701
Jitter	0.04	-0.59 – 0.52	0.890
OffBeat	-0.51	-0.05 – 1.06	0.072
Post:Jitter	-0.69	-0.10 – 1.48	0.084
Post:OffBeat	-0.11	-0.68 – 0.89	0.783
Jitter:OffBeat	-0.43	-1.22 – 0.35	0.276
Post:Jitter:OffBeat	0.44	-1.55 – 0.67	0.433
σ^2	0.35		
Marginal R^2	0.071	Conditional R^2	0.633

P3a Amplitude ∼ Stimulation (Pre/Post) *
Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	1.39	0.70 – 2.08	<**0.001**
Post	-0.14	-0.85– 0.57	0.690
Jitter	-0.24	-0.88 – 0.41	0.468
OffBeat	0.67	0.03 – .321	**0.040**
Post:Jitter	-0.40	-0.51 – 1.32	0.378
Post:OffBeat	-0.71	-1.62 – 0.20	0.126
Jitter:OffBeat	-0.96	-1.87 – 0.05	**0.040**
Post:Jitter:OffBeat	0.71	-0.58 – 1.99	0.278
σ^2	0.47		
Marginal R^2	0.105	Conditional R^2	0.542

Table 2.5: Linear Mixed Effects Model Results – Pre and Post SMA cTBS Stimulation

MMN Amplitude $\sim$ Stimulation (Pre/Post) * Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-0.85	-1.54 – -0.16	**0.016**
Post	0.14	-0.41– 0.69	0.612
Jitter	-0.13	-0.68 – 0.41	0.630
OffBeat	0.65	0.11 – 1.20	**0.019**
Post:Jitter	0.18	-0.59 – 0.95	0.635
Post:OffBeat	-0.29	-1.06 – 0.48	0.456
Jitter:OffBeat	-0.35	-1.12 – 0.42	0.371
Post:Jitter:OffBeat	-0.30	-1.38 – 0.79	0.587
σ^2	0.33		
Marginal R^2	0.064	Conditional R^2	0.685

P3a Amplitude $\sim$ Stimulation (Pre/Post) * Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	1.33	0.57 – 2.09	**0.001**
Post	-0.22	-0.91– 0.47	0.523
Jitter	-0.27	-0.88 – 0.34	0.382
OffBeat	0.46	-0.14 – 1.07	0.132
Post:Jitter	0.45	-0.41 – 1.31	0.296
Post:OffBeat	-0.25	-1.11 – 0.61	0.569
Jitter:OffBeat	-0.25	-1.11 – 0.61	0.569
Post:Jitter:OffBeat	-0.82	-1.68 – 0.04	0.062
σ^2	0.42		
Marginal R^2	0.092	Conditional R^2	0.613

Difference Scores of Pre to Post ERPs Across Stimulation Sites

Finally, we compared the percent difference in ERP amplitude from pre to post cTBS stimulation across each stimulation site, shown in Figures 2.9 and 2.10. The preliminary results for the linear regression model evaluating percent difference after cTBS for each stimulation are shown in Table 2.6. For the change in MMN amplitudes, the effect of condition(jitter) for stimulation(PPC) was statisticallly significant and negative (beta = -1.21, 95% CI = -2.37 – -0.05, p = 0.040), representing an increase (more negative) in MMN for deviants in the jittered condition and PPC stimulation. For the change in P3a amplitudes, the interaction effect of beat(off beat) on condition(jitter) was statistically significant and negative (beta = -1.61, 95% CI -2.86 – -0.36, p = 0.012), representing a decrease in P3a for off beats in the jittered condition. All other differences were non-significant for both MMN and P3a mean amplitudes for PPC stimulation.

[This space intentionally blank for formatting purposes.]

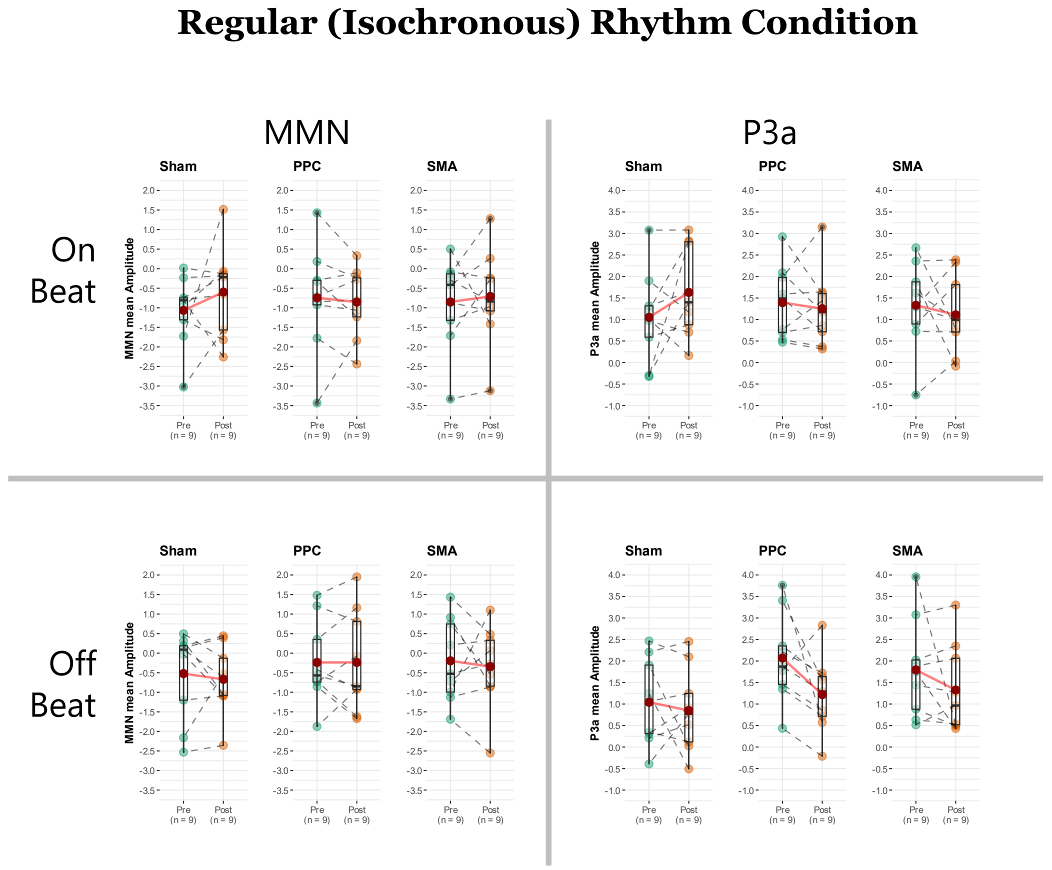
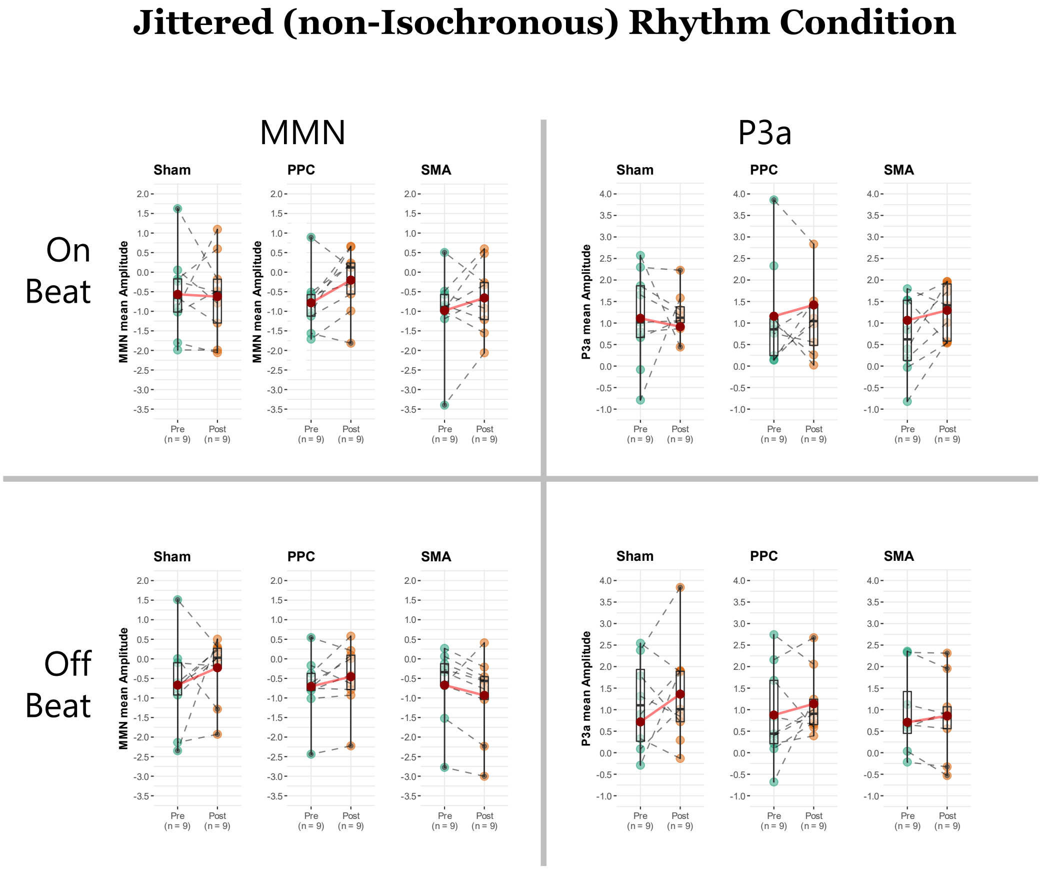


Figure 2.9: Mean ERP amplitude in response to “on” and “off” beat deviants. Lines connect amplitude values from individual participants. Data correspond to the the regular (isochronous) rhythm condition, pre and post cTBS at each stimulation site.



Jittered (non-Isochronous) Rhythm Condition

Figure 2.10: Mean ERP amplitude in response to “on” and “off” beat deviants. Lines connect amplitude values from individual participants. Data correspond to the the jittered (non-isochronous) rhythm condition, pre and post cTBS at each stimulation site.

Table 2.6: Linear Mixed Effects Model Results – Percent Difference after cTBS Stimulation

Difference Score of MMN Amplitude ∼ Target (Sham/PPC/SMA) * Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-0.47	-1.06 – 0.13	0.121
PPC	0.58	-0.24 – 1.40	0.167
SMA	0.33	-0.49 – 1.15	0.429
Jitter	0.52	-0.30 – 1.34	0.207
OffBeat	0.62	-0.20 – 1.44	0.140
PPC:Jitter	-1.21	-2.37 – -0.05	**0.040**
SMA:Jitter	-0.71	-1.87 – 0.45	0.228
PPC:OffBeat	-0.73	-1.89 – 0.43	0.217
SMA:OffBeat	-0.33	-1.49 – 0.83	0.579
Jitter:OffBeat	-1.12	-2.28 – 0.04	0.059
PPC:Jitter:OffBeat	1.56	-0.08 – 3.20	0.063
SMA:Jitter:OffBeat	1.41	-0.23 – 3.05	0.091
σ^2	0.77		
Marginal R^2	0.085	Conditional R^2	0.125

Difference Score of P3a Amplitude ∼ Target (Sham/PPC/SMA) * Regularity (Regular/Jittered) * Beat (On/Off)

Predictors	*Estimates*	*95% Confidence Interval*	*p-value*
(Intercept)	-0.58	-1.28 – 0.12	0.103
PPC	0.72	-0.16 – 1.61	0.109
SMA	0.80	-0.09 – 1.69	0.076
Jitter	0.78	-0.11 – 1.66	0.086
OffBeat	0.77	-0.12 – 1.66	0.088
PPC:Jitter	-1.18	-2.43 – 0.07	0.065
SMA:Jitter	-1.23	-2.48 – 0.03	0.055
PPC:OffBeat	-0.06	-1.32 – 1.19	0.920
SMA:OffBeat	-0.52	-1.78 – 0.73	0.410
Jitter:OffBeat	-1.61	-2.86 – -0.36	**0.012**
PPC:Jitter:OffBeat	0.91	-0.87 – 2.68	0.313
SMA:Jitter:OffBeat	1.45	-0.33 – 3.22	0.109
σ^2	0.90		
Marginal R^2	0.134	Conditional R^2	0.302

2.4 Discussion

Non-human primates, such as the rhesus macaque and chimpanzee, have robust interval timing ability – they can detect detect changes in inter-onset intervals and use interval timing information to plan their actions. However, the beat-based timing ability that allows for flexible perception or movement to a complex rhythm appears to be uniquely human. Previous work has demonstrated that pre-attentive markers of beat perception are present in human ERP responses to complex rhythmic stimuli, but are absent in monkey data (Bouwer et al., 2016; Honing et al., 2018). The results of the current study replicate these effects in human data.

Specifically, there are effects of metrical position on ERP response to deviants in rhythmic sequences. The first aim of this study was to replicate whether there are stronger ERP responses for all the baseline conditions in regular rhythmic sequences vs jittered rhythmic sequences, consistent with beat perception abilities in humans. And to replicate whether there are attenuated ERP responses to off beat deviants, consistent with Bouwer et al 2016, showing that humans are also engaging in sequential learning during even during passive listening. We did observe MMN and P3a components in response to perceptual deviants, and the MMN was lower for off-beat deviants in the regular (isochronous) condition. This means we found a small effect of metrical position on MMN in the predictable, regular (isochronous) rhythm condition. However, we found minimal difference in MMN response based on beat position for the jittered condition, or for P3a in the based on either regularity or beat position. This finding for the jittered (non-isochronous) condition means we did not observe a metrical position effect or a sequential learning effect on the ERPs when the sound onsets are not predictable, despite the sound events themselves being presented in the same order. These overall null results may be due to the fact that we shortened the original protocol from Bouwer et al 2016 down to only 9 minutes and 45 seconds for each rhythmic stimuli, and to the small sample size in this preliminary data covering only nine participants.

The second aim of this study was to extend the findings of previous work by probing the causal role of motor regions implicated in human beat perception. Human beat-based timing perception is thought to rely on tight connections between auditory and motor processes in the brain, such that predictions from motor activity guide sensory response to rhythmic events. We applied TMS to two motor regions: the SMA (implicated in beat-based timing outside of the dorsal auditory stream) and the PPC (implicated in beat-based timing as a hub between motor and auditory cortex within the dorsal auditory stream). If cTBS applied to either SMA or PPC led to an attenuation of these ERP responses, more closely aligned with ERPs in experiments of monkey rhythm perception, then we would have an indication that beat-based timing either inside or outside the dorsal auditory pathway was involved in the underlying oscillatory mechanisms for meter perception or sequential learning in response to these two rhythmic stimuli. Our results with this preliminary data are thus far inconclusive.

Conclusion

To conclude, the aim of this study was to investigate the role of SMA and PPC in beat-based timing perception. We evaluated the neural mechanisms of predictive timing and sequential learning in rhythm cognition by measuring ERPs in response to deviant stimuli in isochronous and jittered rhythmic sequences. As in Bouwer et al., 2016, we show a robust MMN and P3a responses to on beat deviants in isochronous sequences. However, we did not observe a clearly attenuated response to on beat deviants in jittered sequences. Additionally, we observed attentuation of the MMN response in off beat deviants compared to on beat deviants, but only in the regular (isochronous) rhythm condition. This demonstrates that humans are engaging in the process beat perception while passively listening to rhythmic stimuli, but does not clarify to what extent regularity and metrical position affect these neural mechanisms.

We then perturbed these mechanisms using downregulatory TMS. To our knowledge, this represents the first causal investigation – using transcranial magnetic stimulation – into the motor regions and neural mechanisms involved in human beat-perception within (PPC) and outside (SMA) the dorsal auditory stream. Post TMS, we did not observe a clear attenuation of ERP response after stimulation of either SMA or PPC. These results cannot conclusively support either region's role in beat perception activity as implicated in either GAE or the ASAP hypothesis. Significant heterogeneity among participant's ERP response, and response to the cTBS protocol, are both possible factors in this inconclusive data. Analysis of the complete dataset, including comparing any effects of musicianship status, as well as cTBS stimulation thresholds, may provide further evidence regarding the role of enhanced cortical connections between auditory and motor areas in the human brain support beat-based timing perception.

Chapter 3

Measuring Acoustic Social Worlds: Reflections on a Study of Multi-Agent Human Interaction

Chapter 3 provides a conceptual bridging between the first two chapters focused on low-level neural mechanisms in music and rhythm perception, to the final two chapters focused on the rhythms of large group interaction. In this chapter, I review work in musical interaction through the philosophical framework of 4E (embodied, enactive, extended and embedded) cognition. I argue that our implicit understanding of the relationship between the senses and our bodily and motor expertise (a specific view called sensorimotor contingency theory) underpins not only our experience as an individual engaging with the world, but also our experience as an individual musician embedded within larger musical interaction. The chapter ends by briefly summarizing the results of the study presented in chapter five, where these theories inform our empirical investigation of the coordination of members of a large musical ensemble.

In press as:

Proksch, S., Reeves, M., Spivey, M., Balasubramaniam, R. (n.d.) *Measuring Acoustic Social Worlds: Reflections on a Study of Multi-Agent Human Interaction* Eds S. Besser, F. Lysen, & N. Geode, forthcoming with Brill in 2023

Abstract

When individuals are sharing similar behavioral, physiological, or neural states—that is, when individuals' actions, body states, or brain activity are changing together in time—then a collective interpersonal synergy forms (M. A. Riley et al., 2011). Each individual starts to behave together as a member of one large group. Measurement of an interpersonal synergy can be indicative of shared social cognition—of joint participation in co-regulating multiple patterns of activity between two or more agents engaged in a social interaction (De Jaegher, Di Paolo, & Gallagher, 2010). In order to examine the dynamics of multiagent groups of people, and the emergence of these interpersonal synergies, scientists generally measure signals emitted by each individual to detect correlations between these signals.

However, humans are highly capable of tracking complex behavioral dynamics of multiagent groups in our everyday interactions with the world even without access to individual behavioral, physiological, or neural signals from each individual agent. Our everyday human interactions provide us with access to a shared and co-created acoustic social world. Interactions among members of musical ensembles in particular provide useful insight into the co-creation of acoustic social worlds and the emergence of collective synergies. In the context of 4E cognition and the dynamical systems framework, this chapter provides an examination of the measurement of collective synergies from acoustic signals not of individual musicians, but of the acoustic signal co-created by a musical ensemble as a whole, in performance of a musical work that was composed to highlight the emergence of such a collective synergy.

3.1 Introduction

Humans are highly capable of tracking complex behavioral dynamics of crowds in our everyday interactions with the world. Imagine the sounds of a crowded coffee shop. Consider how individuals in that coffee shop might be interacting with each other. There may be some small groups or pairs, and many individuals engaging in small, temporary interactions—but these individuals are not coordinating with every other individual in a cohesive "coffee shop group". They are just a jumble of individuals cohabiting a shared space.

Now, imagine the audience on the floor of a rock concert, cheering or singing along with the artists on stage. Alternatively, imagine the fan section at your favorite sporting event emerging into synchronous chant, or a chorus of resounding "boos". Imagine how individuals in those large crowds might be interacting with each other. As they cheer, or sing, or chant, or boo, they are sharing similar behavioral states—engaging in similar actions. They are likely sharing similar physiological and neural states as well. These crowds are changing together in time. They are behaving—-coordinating—like one large interdependent group. They have formed a synergy.

The differences between this coffee shop crowd and the rock concert audience or the sporting event crowd are trivially easy for you or I to identify. An uncoordinated group of independent individuals, happening to coexist in a shared space versus a coordinated interdependent group of members of a crowd. We can simply hear that these two groups of people sound different. Similarly, we can simply hear when we are engaged and participating in a collective synergy while we are interacting in a large group of people. Or perhaps, in a very large musical ensemble.

As scientists in the lab, we are able to measure one or many signals from every individual in an interacting or non-interacting group. We can measure those signals and, usually, identify which members of the group are sharing the same behavioral state by identifying when their movements synchronize with, or complement, other members of the group. We can measure the pattern of each individual's heartbeats or electrodermal activity (subtle electrical signals conducted by the skin) to identify when individuals are sharing the same physiological state. We can even measure precise patterns of brain activity (EEG, or fMRI) to identify when individuals are sharing the same neural state (Balconi & Fronda, 2020; Misaki et al., 2021; Schirmer, Fairhurst, & Hoehl, 2021). When individuals are sharing similar behavioral, physiological, or neural states—that is, when individuals' actions, body states, or brain activity are changing together in time—then a collective interpersonal synergy forms (M. A. Riley et al., 2011). Each individual starts to behave together as a member of one large group. Measurement of an interpersonal synergy can be indicative of shared social cognition: of joint participation in co-regulating multiple patterns of activity between two or more agents engaged in a social interaction (De Jaegher et al., 2010).

However, as humans going about our everyday lives, we don't carefully measure the individual components of our successful social interactions or the emergence of interpersonal synergies in our joint actions to determine whether we are engaged in a participatory instance of social cognition. We don't carefully monitor our own movement and brain patterns or carefully compare each of these many individual signals that we generate to the many individual signals that our friends, coworkers, or other individuals generate as we are interacting in real time. In fact, we don't even have access to individualised movement, body, and brain data of ourselves and the people with whom we interact on a daily basis.

Scale this up to an interaction between four or five people, or even further to a very large audience attending a concert—and this feat is unmanageable for any single human's sensory system. Often, you might lack even visual access to every member of the group you are interacting in. What you do have access to is a shared and co-created acoustic world (albeit from your own unique point of reference). With this in mind, we turn first to the science of interpersonal coordination, and then to an empirical study evaluating the formation of interpersonal synergy within a set of musicians, who join together in co-creating an acoustic musical world.

3.2 Interpersonal Coordination

The science of interpersonal coordination has made advances in describing how individuals interact as part of a dyad or a large group by evaluating a multitude of movement, body, and brain signals from each individual in comparison with each other individual.

But what if scientists don't have access to the vast array of recording devices they rely upon in their lab? What if a scientist wants to study how real groups interact in the wild? Can science identify these same differences in an uncoordinated group of independent individuals versus a coordinated, interacting and interdependent crowd? This is the question we asked in a recent study of multi-agent interaction within a musical ensemble. These musicians performed a piece that was specifically composed so that the musicians first create uncoordinated noise for a period of time on each of their instruments, before joining together into a coordinated joint musical performance. Proksch et al. (2021) wanted to understand how the musicians were changing their acoustic behavior in time, either independently or interdependently, in the two different musical interactions dictated by their musical score. With the understanding that individuals often do not have unobstructed visual access to every member of an interacting group (let alone physiological or neural data), we restricted our dynamical systems analysis to a single measurement of the shared and co-created behavior generated by the musicians—a raw audio file of the ensemble's acoustic behavior.

The dynamical systems framework in cognitive science allows for the study of the formation of interpersonal synergies. An interpersonal synergy occurs when the movement dynamics of one individual become causally coupled to the movement dynamics of another individual (M. A. Riley et al., 2011). This means that the actions of cognitive agents constrain each other, interacting as a single coupled unit. Interpersonal synergies can arise from simple interactions, such as walking through the park engaged in a conversation and finding one has begun walking in step with your conversation partner (Atherton, Sebanz, & Cross, 2019). Subtler interpersonal synergies can arise in conversation when standing still. Even if one cannot see their conversation partner, the mere act of interacting through conversation serves to constrain subtle sway patterns of body movement, such that body movement is distinctly coupled to the movement of the unseen partner (Shockley, Santana, & Fowler, 2003).

Perhaps more immediately observable, however, are the interpersonal synergies which we see and hear in musical interactions. Where conversation partners might incidentally fall into step or sway together in time, a pair or a group of musicians co-creating a musical performance are intentionally coordinating their acoustic behavior. It's important to note here, that coordinating acoustic behavior in order to engage in a successful musical interaction often involves musicians moving their bodies differently from their musical partners. A trombone player will make different movements than the string bass player, and a trumpet player or pianist will make different movements and perhaps even play more notes in the same amount of time compared to the trombonist and bassist. But together, these differing movement dynamics from each musician join to co-create the same shared acoustic social world. The acoustic output of each musician constrains the acoustic output, and motor behavior, of each other musician in the ensemble. In fact, if the low voices (the trombone and the bassist) were to play a continuous drone, one single chord for a prolonged time, then the duet that the trumpet player and pianist improvise together may result in a different 'performance narrative' than if the low voices provided a rhythmic bass line. In a study involving duets performed by pairs of skilled pianists, improvising over the unstructured 'drone' backing track resulted in increased movement coordination between the two pianists compared to improvisation over the rhythmic bass line (Walton et al., 2018). Specifically, pianists repeated their improvisation partner's note combinations and head movements in longer sequences when improvising over the drone backing track. Further, listeners rated this performance as more 'harmonious' than the improvisation over the structured, rhythmic bass line, with listeners giving higher 'harmonious' ratings when the musicians repeated each other's note combinations for longer sequences of time (Walton et al., 2018).

Experimental setups are typically designed to identify interpersonal synergies by correlating one or more of the movement/body/neural signals from each member of the interacting or non-interacting group. But in principle—once an interpersonal synergy is formed—it should be possible to analyze group behavioral dynamics from one single signal measured from that system. This is due to two factors—dimensional compression, and reciprocal compensation (M. A. Riley et al., 2011). Dimensional compression within a synergy occurs when the movement of many potentially independent elements (such as the movement of two independent pairs of legs on two independent walking individuals) become coupled so that they move in time together (the two pairs of legs begin to walk in step, as one interdependent walking dyad). Reciprocal compensation, also termed mutual adaptation, describes the ability of movement in one element of a synergy to react to, or adapt to, the movement of another element of the synergy (one member of the walking dyad can adjust their walking speed to ensure they are in step with their walking partner; Riley et al., 2011). The behavior of the musicians in the improvising piano duets we visited earlier exhibited these features of dimensional compression and reciprocal compensation. The movement of two independent pairs of hands, and two independent heads, on two individual musicians became coupled so that they created music in time together, and the musical behavior of each musician reacted to, or adapted to, the musical behavior of the other musician. The listeners, who rated this musical performance, were able to extract an acoustic signal from that system and attune to differences in how the two duetting pianists coordinated their sound and movement (Walton et al., 2018). If these listeners were able to attune to these differences in coordination in two forms of coordinated music making (improvised duets over two different backing tracks) based on a single acoustic signal—the raw audio of the music performance itself—then perhaps this same feat can be scaled up to a multi-agent interaction of a much larger musical ensemble. And if the motor and acoustic behavior of individuals within a much larger musical ensemble are functioning together in time, so as to have the features of dimensional compression and reciprocal compensations necessary to form an interpersonal synergy, then it should be possible to detect that synergy from something as sparse as a raw audio file.

3.3 An Empirical Study of Multi-Agent Musical Interaction

We investigated the coordination dynamics of a performance of "Welcome to the Imagination World", composed by Daisuke Shimizu (Shimizu, 2016) and performed by the Inagakuen Wind Orchestra (Proksch et al. (2021))[1]. Specifically, we evaluated the acoustic behavior of this musical ensemble using methods from dynamical systems theory of phase space reconstruction and recurrence quantification analysis. These time series analysis methods allow researchers to detect two features of interpersonal synergies discussed above—dimensional compression and reciprocal compensation—and to measure patterns of this synergistic behavior over time. In this case, we were somewhat playing the role of the listeners of the improvising pianists. The difference was, instead of asking whether one could hear differences in coordination, the question was whether one might empirically measure differences in coordination using those dynamical systems tools. And importantly, can these differences in coordination be measured using only the raw audio signal

---

[1] https://www.youtube.com/watch?v=-wJ9ZsgO3QI

of the musical performance, without access to individual recordings of each musicians' acoustic output.

The musical performance was divided into two main coordination categories: uncoordinated and coordinated. This uncoordination was in fact a specific feature of the musical composition itself. Shimizu composed this piece to reflect the "arrival and development of a simple fanfare motif into an accomplished work" (windrep.org), beginning with "random ad lib music...free of tempo and as expressive as possible" until the musicians invite the conductor on stage as the horn, tenor, and brass instruments unify into a majestic introduction' (windrep.org). These descriptions are from program notes describing "Welcome to the Imagination World". Listening to a performance of this work, one can easily hear the difference between the uncoordinated improvisations of individual musicians on stage and the coordinated, collective interaction of the musicians as they co-create 'an accomplished work'. Importantly, however, we were also able to measure those differences in coordination dynamics from the raw audio signal.

The results from the time series analysis make clear that there is a detectable difference between the uncoordinated and coordinated portions of the performance. They tell us that when the musicians began to coordinate their actions, such that the actions of each musician became interdependent on the action of each of the other musicians in the ensemble, they formed a single complex system—a collective interpersonal synergy.

3.4 Discussion

What are the implications of measuring these differences in coordination dynamics between the acoustic behavior of non-interacting uncoordinated musicians gathered on a stage versus the acoustic behavior of the same musicians when interacting and coordinated as a cohesive musical ensemble?

The current pragmatic turn in cognitive science toward action-oriented views of cognition (c.f. Engel, Friston & Kragic, 2016) provides a useful explanatory viewpoint for discussing coordination as it relates to multi-agent musical interaction. Specifically, we can interpret the collective coordination which emerged in this musical ensemble in terms of sensorimotor contingency theory under the cognitive framework of enactivism. We argue that the skilled coordination of these musicians engaging in joint musical action is grounded in (implicit or explicit) knowledge of sensorimotor contingencies supporting music perception and production. These shared sensorimotor contingencies enable an interacting multi-agent group of musicians to co-create a shared, acoustic social world—forming a single complex system—as the interdependent actions of individual musicians give rise to emergent dynamics of an interacting musical ensemble.

Enactive sensorimotor contingency theory is a theory of perception, which describes perception as a process which is guided by action, emphasizing a “pre-conceptual, pre-linguistic form of understanding related to bodily and motor expertise” (Matyja & Schiavio, 2013). Developed originally as an explanation for visual perception (O’Regan & Noë, 2001; Noë, Noë, et al., 2004), the classic example of sensorimotor contingency theory is ‘seeing the whole tomato’. When we see a tomato, we don’t just see a two-dimensional gradient of colors and edges, but in some sense, we see the ‘whole’ tomato. Our awareness of the back of the tomato arises from our bodily knowledge of a repertoire of motor actions, the sensorimotor contingencies (SMCs), necessary in perceiving tomatoes: we know that if we were to perform a certain action (turn the tomato around), that we would see the back of a tomato. Enactivist accounts of music cognition place the perception of music in (implicit or explicit) bodily knowledge of a repertoire of motor actions and their effect on associated sensory stimulation, or knowledge of sensorimotor contingencies (Matyja & Schiavio, 2013). Rather than passive listeners, simply receiving a barrage of acoustic stimuli and later appraising it as musical (as in a traditional, cognitivist account of music perception), we perceive music through skilled action (Maes, Leman, Palmer, & Wanderley, 2014; Maes, 2016). The music listener learns to ‘manipulate’ the barrage of acoustic stimuli she hears through active (attentive) listening and skillful engagement with the musical environment (Krueger, 2009, 2013). Musical training or experience enhances her knowledge of the sensorimotor contingencies involved in producing music, which enables her to selectively attend to increasingly more salient musical features when perceiving music. Knowledge of SMCs involves not only knowledge of what sound can be heard given a certain action, but also what action most likely caused or will cause a certain sound. This bidirectional association between action and perception enables individuals to plan and respond to their own action, and also to predict and coordinate with others through joint musical action—as listeners and players at the same time.

A series of studies by Drost et al. (2005, 2007) demonstrate that individuals with musical training were more susceptible to making mistakes in a forced production task due to incongruencies between visually and auditorily presented chords, and that the effect was stronger when the auditory stimulus presented was in the timbre of their own instrument. These studies indicate that musical training leads to more precise sensorimotor representations of the action necessary to produce a sensory stimulus (the heard chord). A number of piano timing experiments demonstrated that pianists (ignorant of the task condition, and told that they were performing with a live partner) were better able to play in time with recordings of themselves than of other musicians

(Keller, Knoblich, & Repp, 2007) or with others who were matched in terms of preferred performance tempo (Loehr, Large, & Palmer, 2011). Each of these studies indicate that higher knowledge of SMCs enhances the participant's ability to be in time in a music production task by enhancing the participant's ability to plan and coordinate with a partner in joint musical activity.[2]

Rhythmic interaction in naturalistic music making, such as musical ensemble performance, relies on extending these simple sensorimotor contingencies—knowledge of the sound your instrument will make when you perform an action; knowledge of the sounds you will hear based on the preferred tempo at which you play music—to more complex SMCs which take into account the dynamics of two or more interacting agents, such as knowing what musical phrase you will hear from your band mates in a jam session after you have each taken a certain set of musical actions. Humans excel at the precise timing and coordination of motor and acoustic output from multiple musicians in part because they excel at a skill called entrainment, which is where we are headed next.

Successful coordination within and between human individuals in music making may reflect a greater (implicit or explicit) knowledge of the sensorimotor contingencies involved in perceiving and producing musical events. Take our pianists for example, they are better able to synchronize with recordings of themselves (Keller et al., 2007) or with others who are matched in preferred performance tempo (Loehr et al., 2011). This is because the pianist (unknowingly) playing with a recording of herself has a very strong implicit knowledge of what actions it would take for her to produce the sound she hears from the recording. This strong knowledge of SMCs makes it easier for her to predict when and what she will hear, and enhances the strength of entrainment between musician and recording. It is thus easier for the pianist to form an interpersonal synergy alongside their own pre-recorded musical activity. The pianist who is playing with another who prefers similar tempi has a similarly strong knowledge of SMCs involved in producing the sounds they hear at their preferred tempo, enhancing the strength of entrainment at that tempo, and enabling the emergence of a tight interpersonal synergy.

Entrainment, or a specific form of coordination referring roughly to the ability to synchronize or to be together in time with one or more individuals, has been taken to "relate phenomenologically to a sense of social belonging"

---

[2] The original authors interpretation of these experiments and results was taken as support for the role of cognitive representations of the actions of self and other in musical activity. Thus, by their account, musicians are cognitively simulating the movement and production of the other musician. Under moderate accounts of SMC, these representations may be thought to include knowledge of SMCs, however under more radical accounts of SMC (with no role for representations) these studies may be taken to support the role of bodily knowledge/memory of the SMCs involved in music production and response.

and has been conceived as one explanation for group cohesion and bonding that emerges from joint activities such as music making (Clayton, Sager, & Will, 2005). Our human ability to entrain with others ranges from the subconscious synchrony of repetitive motions (i.e. happening to walk in step with another) to the synchronization of intentional temporal events such as synchronizing melody and harmony in joint musical interaction. In temporal rhythmic processing, it is the interaction of the body, brain, and environment which results in the emergent phenomenon of sensorimotor and neural entrainment (J. M. Ross & Balasubramaniam, 2014). Interpersonal synchronous movement between two or more individuals may be further linked through synchrony of neural oscillations across individuals (Novembre, Knoblich, Dunne, & Keller, 2017) and has been found to occur in naturalistic social interactions among affiliative partners (Kinreich, Djalovski, Kraus, Louzoun, & Feldman, 2017). Even when referring in part to neural phenomena, such multi-scale and multi-level coordination patterns likely relate to our phenomenological experience of being part of an extended social and cultural environment. Kirchoff and Kiverstein describe this feeling as “phenomenal attunement— the feeling of being at home in a familiar culturally constructed environment” (Kirchhoff & Kiverstein, 2020). The interactions of a musical ensemble, specifically when the ensemble is made up of a group of musicians who have engaged in repeated rehearsals and joint musical interactions together, provides an ideally structured social and musical environment for that ensemble to exhibit an extended cognition, if not an extended conscious mind (Spivey, 2020; Kirchhoff & Kiverstein, 2020).

In perhaps a less enactivist light, shared predictive models of sensorimotor contingencies developed during and as a result of group music making may give rise to group identity in a similar fashion to the “predictive perception of sensorimotor contingencies” which are proposed to underlie a sense of self (Seth, 2014). While radical enactivism maintains a strictly antirepresentational character of the nature of cognitive processes (c.f. Hutto & Myin, 2012), we do not take a stance in this debate in this chapter. Rather, we argue that enactive SMC theory grounds and enhances aspects of coordination in joint musical action in terms of bodily and environmental states, regardless as to whether these states are represented in the brain as models or wholly constituted by the bodily/environmental states themselves.

Individuals engaged in joint music making often join together in larger groups than these piano duets, ranging from a four-person quartet to a large chorus, orchestra, or even a stadium full of concert goers singing along with their favorite band on stage. Nevertheless, the shared sensorimotor contingencies among multi-agent groups of interacting musicians enable them to co-create a shared, acoustic social world. In doing so, they form collective interpersonal synergies, allowing the interdependent actions of individual musicians to give rise to the emergent dynamics of an interacting musical ensemble. By examining these synergies in the context of 4E cognition (Newen, De Bruin, & Gallagher, 2018), we can see them as emerging from groups of agents who are embodied and enactive, as well as embedded in an environment, thus making their cognition extended across many interacting elements. That is, when the people and their instruments are well coordinated by virtue of their shared and co-created acoustic social world, they form one complex system that, by itself, bears a substantial statistical similarity to a mind.

Chapter 4

Coordination Dynamics of Multi-Agent Interaction in a Musical Ensemble

Chapter 4 is a published empirical case study, applying nonlinear statistical analysis techniques to investigate the interaction of musicians within a musical ensemble. With the theoretical background provided in chapter 3, I demonstrate that we can measure one of the many possible sources of sensory and behavioral information, and characterize two different modes of coordination. Specifically, nonlinear analyses applied to the acoustic signal created by these musicians can be used to differentiate when musicians are engaged in individual - independent - musicianship, and coordinated - collective - musical ensemble performance. This chapter concludes with a call to extend these methods to study other forms of large group social interaction 'in the wild', anticipating the study of crowd sound interaction presented in chapter five.

Published as:

Proksch, S., Reeves, M., Spivey, M., Balasubramaniam, R. (2022) Coordination Dynamics of Multi-Agent Interaction in a Musical Ensemble. *Nature Scientific Reports*

Abstract

Humans interact with other humans at a variety of timescales and in a variety of social contexts. We exhibit patterns of coordination that may differ depending on whether we are genuinely interacting as part of a coordinated group of individuals vs merely co-existing within the same physical space.

Moreover, the local coordination dynamics of an interacting pair of individuals in an otherwise non-interacting group may spread, propagating change in the global coordination dynamics and interaction of an entire crowd. Dynamical systems analyses, such as Recurrence Quantification Analysis (RQA), can shed light on some of the underlying coordination dynamics of multi-agent human interaction. We used RQA to examine the coordination dynamics of a performance of "Welcome to the Imagination World", composed for wind orchestra. This performance enacts a real-life simulation of the transition from uncoordinated, non-interacting individuals to a coordinated, interacting multi-agent group. Unlike previous studies of social interaction in musical performance which rely on different aspects of video and/or acoustic data recorded from each individual, this project analyzes group-level coordination patterns solely from the group-level acoustic data of an audio recording of the performance. Recurrence and stability measures extracted from the audio recording increased when musicians coordinated as an interacting group. Variability in these measures also increased, indicating that the interacting ensemble of musicians were able to explore a greater variety of behavior than when they performed as non-interacting individuals. As an orchestrated (non-emergent) example of coordination, we believe these analyses provide an indication of approximate expected distributions for recurrence patterns that may be measurable before and after truly emergent coordination.

Keywords: Coordination Dynamics, Social Interaction, Emergence, Dynamical Systems, Music

4.1 Introduction

Science has looked to art for inspiration in explaining human cognition. Music, in particular, has aided scientists exploring human engagement with the world, from emotional experience (Meyer, 1956; Huron, 2008; Juslin & Västfjäll, 2008) to social interaction (D'Ausilio, Novembre, Fadiga, & Keller, 2015; Walton, Richardson, Langland-Hassan, & Chemero, 2015; Chang, Livingstone, Bosnyak, & Trainor, 2017; Walton et al., 2018; Chang, Kragness, Livingstone, Bosnyak, & Trainor, 2019). Music provides an ideal model system of human social interaction—balancing ecological validity of the interaction and environment with experimental control (D'Ausilio et al., 2015).

Consider the initiation of the slow clap by one, then two, then four people, before breaking into full audience applause or the first musician in a flash mob initiating a flow of musicians and audience members engaging in shared music making. The truly emergent sound of audience applause, and the script-guided pseudo-emergent sound of a musical flash mob each provide examples of acoustic behavioral patterns showcasing the transition from individual behavior to multi-agent interaction. Studying the patterns which arise from pseudo-emergent coordination aided by a musical script can shed light on some of the coordination dynamics which underlie truly emergent multi-agent human interaction.

Transitions from disorder to order are exhibited by a variety of animals ranging from locusts marching (Buhl et al., 2006) to birds flocking (Cavagna et al., 2010) to humans clapping (Néda, Ravasz, Bréchet, Vicsek, & Barabási, 2000). The patterns and conditions for this emergent coordination between individuals has been a subject of laboratory study for decades. Spontaneous, or emergent, patterns of entrainment are measured between individuals by analyzing video and motion capture from interacting dyads swinging pendulums (Schmidt & O'Brien, 1997) or rocking in rocking chairs (M. J. Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). Recent work has evaluated motor coordination dynamics of naturalistic interactions such as interactive problem solving (Abney, Paxton, Dale, & Kello, 2015; Nalepka, Kallen, Chemero, Saltzman, & Richardson, 2017), naturalistic conversation between individuals (Paxton & Dale, 2017; D. C. Richardson & Dale, 2005), speed-dating partners (Chang et al., 2020), and motor and acoustic coordination of performing musicians in duets (Walton et al., 2015, 2018) and larger ensembles (Chang et al., 2017, 2019). In these interactions, behavioral output of each interacting individual was measured and analyzed for meaningful correlations *between individuals*. An investigation of emergent synchrony in audience applause explicitly measured acoustic output of the group (Néda et al., 2000). However, the motor behavioral patterns were still measured from the *local* behavior of *individual* audience members in order to evaluate correlation with the global signal of the audience. In each of these studies, it has been possible to obtain clear measurements of individual behavior to examine the emergent coordination dynamics of multi-agent interaction and social self-organization. However, if that multi-agent group has indeed self-organized into a complex system, then the interdependence between the agents' functions should make it possible to detect that coordination from almost any time series emitted from that system, using state-space reconstruction (Takens, 1981; Vlachos & Kugiumtzis, 2010). Thus, if obtaining movement or acoustic measurements at the individual and local level is not feasible or practical with group behaviors in the wild, then global-level measurements should suffice. One of the simplest global measures to employ is an acoustic recording taken from a well-placed microphone. Increasingly, naturalistic recordings from e.g. Youtube are being used in human behavioral research where specific recording equipment is unknown, and likely does not contain visual or auditory signals mapped to individual-level behavior in group settings. Benefits of these sorts of recordings include heightened ecological validity and real world behavior. Already, Alviar et al 2020 and Kello et al 2017 analysed coordination between sound and movement, and multiscale structure in orchestral music, jazz, TED talks, and even animal vocalizations through a collection of videos found on youtube (and other corpora) with ostensibly varying recording setups, number and type of microphones, etc (Alviar, Dale, Dewitt, & Kello, 2020; Kello, Bella, Médé, & Balasubramaniam, 2017).

This paper shows that it is possible to describe coordination patterns of multi-agent interaction by analyzing a time series extracted solely from group-level audio data. Rather than analyzing the individual-level behavior of interacting agents, we used nonlinear methods from the dynamical systems framework on *group-level acoustic data*. We analyzed global patterns of coordination in a musical performance of "Welcome to the Imagination World" (Shimizu, 2016). This performance enacts a physical simulation of an orchestrated (non-emergent) transition from uncoordinated to coordinated interaction. We used Recurrence Quantification Analysis (which relies on state-space reconstruction) to investigate patterns of coordination from the audio signal of a performance of this work(InagakuenWindOrchestra, 2009). Although this is analysis is being applied on a single recording, we believe recurrence measures of this orchestrated musical performance provide an indication of the possible expected distributions for recurrence patterns that may be observable before and after spontaneous emergent coordination. Moreover, analysis of the transition itself from uncoordinated to coordinated behavior may provide insight into the trade-off between the playful enjoyment of novelty and the rigor of predictive success (Kiverstein, Miller, & Rietveld, 2019). Finally, we discuss applications to other examples of real-world multi-agent human interactions, such as multi-agent interaction at sporting events or of individuals coordinating in a protest.

4.2 Human Behavior as a Complex Dynamical System

Principles of Complex Dynamical Systems

A canonical example of a simple dynamical system is the pendulum clock. A pendulum is a mechanical device—e.g. a fixed weight on a string—which oscillates isochronously around a central point, meaning that swings in both directions take equal amounts of time. The consistent rate of oscillation made the pendulum clock an ideal time-keeper following its invention in the seventeenth century by Cristian Huygens. The pendulum clock and the metronome are examples of simple dynamical systems. The state (location) of the pendulum at any given time is determined by the trajectory of the pendulum over historical time. The oscillatory behavior of a pendulum can be explained by a system of differential equations.

What is relevant here is the behavior that emerges among two or more pendulums placed on a shared surface. Huygens observed that two pendulums hanging from a single beam will spontaneously—or emergently—synchronize their behavior, swinging simultaneously in anti-phase with one another. Multiple metronomes placed atop a platform balanced on two cylinders will also demonstrate emergent synchronized behavior (Pantaleone, 2002; Francke, Pogromsky, & Nijmeijer, 2020). A metronome is a special type of pendulum, which clicks at isochronous intervals to aid time-keeping for musicians. Metronomes feature a fixed weight at the base of a rod, in addition to a moveable weight which slides along the top of the rod to adjust the speed of the metronome oscillations, and thus the speed (tempo) of the audible metronome clicks. If set at the same tempo, the oscillating pendulum and audible clicks of multiple metronomes will begin to synchronize both in-phase and anti-phase with one another (Francke et al., 2020). What begins as multiple individual metronome clicks will transition to clicks occurring simultaneously, as globally isochronous acoustic events. For both the pendulum clocks and the clicking metronomes, the local behavior of each individual pendulum is coupled to the behavior of the surrounding pendulums due to their behavior within a shared context — in this case a physical connection via a single beam or a single platform. The local behavior of each individual metronome or pendulum (each oscillating at approximately the same frequency but different phase) eventually self-organizes into *emergent* global patterns of synchronized behavior.

We observe similar patterns of behavioral synchronization in multi-agent human interaction. The shared context for multi-agent human behavior need not be a physical connection like the metronomes' shared platform. Rather, the shared context mediating emergent global patterns of human behavior is the *interaction* itself. When local behavior of individual human agents becomes coupled to the behavior of surrounding agents—via the shared context of interaction—the agents self-organize to exhibit emergent global patterns of coordinated behavior. We can detect this emergent coordination of multi-agent human groups from their acoustic behavior over time, just as we can detect the emergent coordination of metronomes from their acoustic output over time.

Principles of Recurrence Quantification Analysis

Multi-agent human behavior, such as in crowds or musical ensembles, can be considered as complex dynamical systems, where complex global patterns of behavior emerge as a result of the self-organization of individual agents over time acting according to simple local rules. The behavior of such a dynamical system can be visualized in recurrence plots. These recurrence plots display the system's trajectory through a phase-space, depicting when that trajectory revisits locations within that phase-space at each moment in time. Recurrence quantification analysis is used to describe the complexity of a system over time by analyzing small-scale structures in the recurrence plot.

There are a few key concepts underlying the generation of recurrence plots from time series data. In a recurrence plot, the time series data will sit on a plot with axes of time by time. A point (i, j) is plotted if the value at time i and time j are sufficiently similar — that is, *recurrent* — within a specified neighborhood size of the N-dimensional state-space (Balasubramaniam, Riley, & Turvey, 2000). The state-space of a dynamical system is the vector of possible combinations of states in some number of observable and unobservable dimensions. In order to determine which points are recurrent, it is necessary to reconstruct the higher dimensional phase space of the system. *State-space reconstruction* is done by embedding the original time series against a time lagged copy of itself (M. Riley, Balasubramaniam, & Turvey, 1999). Each time lagged copy is an additional *embedding dimension* within the state-space (Stephen, Dixon, & Isenhower, 2009). Takens' theorem (1981) shows that the coupling of activity between dimensions preserves the information dynamics of the system as a whole in any single dimension. Put another way, because the subcomponents of a complex dynamical system are intrinsically interdependent, a measurement taken from any one observable subcomponent encodes information from every other (potentially unobservable) subcomponent in the system. Thus, reconstruction of the N-dimensional state-space from a single measured time series allows us to infer the topological dynamics of a multivariate system because the influence of higher dimensional dynamics is encoded in the measured dimension (Marwan & Webber, 2015)

The logic of state-space reconstruction and embedding within higher dimensions is important for evaluating the dynamics of a natural complex system. The dynamics of natural systems such as crowd behavior, sounds within a piece of music, or even weather patterns, contain N possible state variables as well as N possible combinations of nonlinear and bidirectional interactions. State-space reconstruction allows us to infer these unmeasured or unobservable higher order dynamical variables from a single measured variable, in order to evaluate the characteristic dynamics of a system's behavior over time. For a weather system, we might measure the flow of high- and low-pressure systems to evaluate the transitions from stable, 'good' weather to instability that precedes a storm. In a crowd or a musical performance, we might measure movement or acoustic signal to investigate the higher order dynamics of transitions between periods of instability and incoherence to periods of stability and coordination. Here, we take a simple global measure of acoustic signal recorded from a musical performance. This performance demonstrates an orchestrated (non-emergent) transition from uncoordinated to coordinated interaction (described in more detail below). We use RQA to investigate the patterns of coordination in the audio signal as represented in small-scale structures within recurrence plots. We focus on five key measures: recurrence rate, determinism, entropy, average diagonal length, and laminarity.

4.3 Music and multi-agent human intearction: A model from acoustic data

Musical ensembles as models of human social interaction

Music provides an ideal model system of human social interaction by providing a balance between ecological validity of the social interaction and experimental control(p111) (D'Ausilio et al., 2015).

Analysis of social interactions in musical performance are aided by a "script-like description of the interaction" via the musical score that can be manipulated or referenced by researchers examining the behavioral dynamics of the interaction (p112) (D'Ausilio et al., 2015). Applying methods of Granger-causality to motion capture data of individual musicians within a performing string quartet can be used to investigate how predictive the history of behavior of one musician is for the future behavior of another. The bodily sway dynamics of these interacting musicians carries Granger-causal information about leader and follower behavior of each musician (Chang et al., 2017), and Granger-coupling of bodily sway also carries information about the joint emotional expression and perceived emotional intensity of a musical performance (Chang et al., 2019). Even without reference to a strict musical score, studies of musical interaction have provided insight into how we anticipate and adapt to the behavior of other individuals. Nonlinear analysis techniques have revealed spontaneous self-organizing patterns of coordination across a variety of timescales during joint musical improvisation without a strict score. In a series of experiments analyzing interactions between improvising musicians, Walton et al. describe how behavior produced and received from both the kinesthetic and sonic domain serves to influence and constrain mutual improvisers from the lens of complex dynamical systems (Walton et al., 2015, 2018) . Using cross-wavelet spectral analysis and Recurrence Quantification Analysis, Walton et al. describe how mutual behavioral constraints enable an improvising ensemble to produce more complex patterns than any individual would otherwise(Walton et al., 2015, 2018). This mutual interaction establishes a single synergetic system at the level of the improvising group, rather than a set of individuals behaving as single agents.

Rather than analyzing the individual-level behavior of interacting musicians, we apply dynamical systems analysis — specifically RQA — on *group-level acoustic data* to analyze global patterns of coordination in a musical performance. We chose to analyze this performance because it enacts a phenomenological simulation (described below) of the transition from the uncoordinated behavior of individuals to coordinated group behavior that mimics naturalistic multi-agent human interaction.

The composition “Welcome to the Imagination World” composed by Daisuke Shimizu for wind orchestra serves as the model system for multi-agent human interaction. Specifically, the interaction of interest is the shift in dynamics from an uncoordinated, incidental collection of musicians, to a coordinated, interacting ensemble. This transition from uncoordinated, to coordinated interaction is evident in the phenomenological experience of attending (or indeed, performing) this piece of music. The audience will note that, at first, the musicians on stage have no conductor. They sound and look like they are each playing their individual warm up routine. This is because, in fact, the musicians’ score tells them to play at random. The composer wanted the sound to be aleatoric, or to occur by chance without being strictly composed. This uncoordinated soundscape continues until a melodic pattern starts to emerge from a few of the musicians, still in the absence of a conductor, and still not appearing to be coordinating with the other performers. Next, the “conductor walks on stage [as] the horn, tenor and bass instruments unify into a majestic introduction”, according to program notes from the composer (windrep.org). This marks the transition from the uncoordinated actions of individual musicians to the coordinated ensemble musicianship the audience expects. The remaining musical score is composed to dictate the acoustic interaction of the musicians on stage. Thus, the rest of the performance demonstrates the coordinated interactions of an interdependent complex system: the multi-agent musical ensemble.

4.4 Results

Recurrence Quantification Analysis

Figures 4.1 and 4.2 show recurrence plots generated from the time series data of the recording. The recurrence plots visualize the characteristic patterns of recurrence which are then quantified through recurrence quantification analysis. Figure 4.1 displays recurrence plots for the first and last 30 seconds of each Coordination Category (Uncoordinated or Coordinated). Darker segments of the recurrence plots indicate the presence of more recurrent data points. Vertical and horizontal lines indicate periods of stability in the system, where one state was visited for a period of up to a few seconds at a time.

The recurrence plots in Figure 4.2 are representative 5-second samples drawn from each 30 second sample in Figure 1. These shorter samples are labeled as the associated Performance Event the 5-second sample is drawn from. Note that the 'Introduction' section falls in the middle of the performance, marking the shift from the Uncoordinated to Coordinated interaction among the musicians, and thus the order of Performance Events is: 1. Aleatoric, 2. Transition, 3. Introduction, 4. Finale. Thirty second samples from the start and end of each Coordination Category were chosen to maintain balanced samples from each category for computing inferential statistics, and because independent raters noted a clear Transition section in the final thirty seconds of the uncoordinated performance.

Increased presence of diagonal lines and vertical structures in the Introduction and Finale sections indicate increasingly coordinated interaction among the musicians. The diagonal striping in the Introduction indicates some periodicity in the signal, similar to the periodicity of a sine wave, and results in this case from a(n almost) unison chord during those few seconds. The Aleatoric and Transition plots at the top more closely resemble white noise, with fewer recurrent points, shorter diagonals, and less apparent vertical structures.

Recurrence Quantification Analysis quantifies the qualitative patterns observed in these recurrence plots. To describe the behavior of our model multiagent system, five RQA measures were evaluated. The first RQA measures evaluated are common measures of recurrence: a) patterns that repeat over time (Recurrence, the percentage of recurrent points on the recurrence plot); b) behaviors that belong to a longer sequence of behavior (Determinism, or the percentage of points that fall on any diagonal line in the recurrence plot); c) the amount of disorder there is in these sequences (Entropy, or the variability in lengths of these diagonal lines). Additional patterns of stability in the system's behavior were measured by examining d) clusters of behavior (Laminarity, or the percentage of points that fall on a vertical line in the

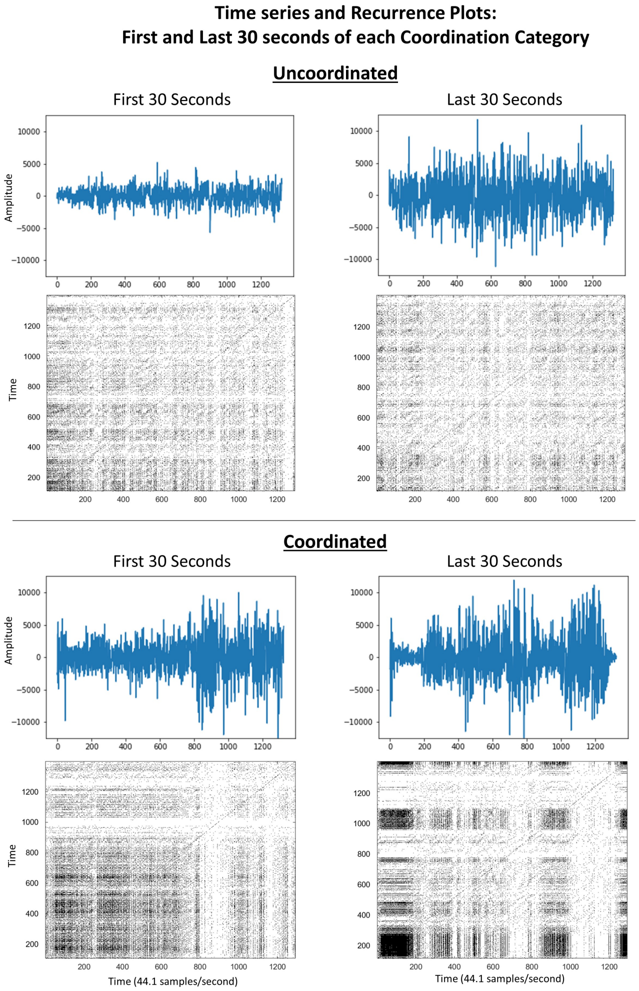


Figure 4.1: Time Series and Recurrence Plots for the first and last 30 seconds of each Coordination Category (Uncoordinated or Coordinated). Darker segments of the recurrence plots indicate the presence of more recurrent data points. Vertical and horizontal lines indicate periods of stability in the system, where one state was visited for a period of up to a few seconds at a time. Note: 9 seconds of audience applause during the Introduction Performance Event were not analyzed, and are subsequently excluded from all data visualizations.

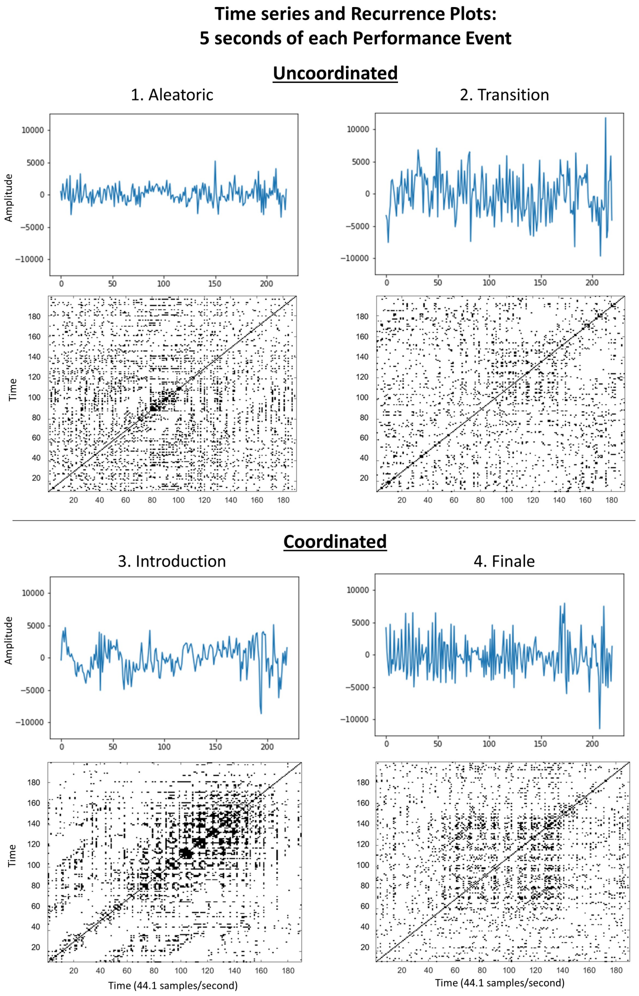


Figure 4.2: Time Series and Recurrence Plots for representative 5-second samples drawn from each 30-second sample in Figure 1, and labeled as the associated Performance Event the 5-second sample is drawn from. Uncoordinated Aleatoric and Transition plots more closely resemble white noise, with fewer recurrent points, shorter diagonals, and less apparent vertical structures. Increased presence of diagonal lines and vertical structures in Coordinated Introduction and Finale plots indicate increasingly coordinated interaction among the musicians. Note: 9 seconds of audience applause during the Introduction Performance Event were not analyzed, and are subsequently excluded from all data visualizations.

recurrence plot), and e) the average length of time our multi-agent system stays in one behavioral pattern (average diagonal length, the average length of diagonal lines). Average diagonal line length is a measure related to determinism. Longer average diagonal lines reflect the stability of a system by indicating longer, more continuous states. Similarly, higher laminarity shows the rigidity, or ‘stickiness’ of a system that stays in one or more states of a behavior for a length of time (Davis, Pinto, & Kiefer, 2017).

Higher values and increased variability were evident for most RQA measures for Performance Events within the Coordinated sections of the performance vs Uncoordinated sections (Figure 4.3). However, distribution plots do not readily visualize different trajectories of behavior over time. Varying trajectories of each RQA measure for Coordinated vs Uncoordinated sections of the performance is evident in the serial plots in Figure 4.4. The majority of RQA values hover around a single value over time during the Uncoordinated Performance Events (Aleatoric and Transition), indicating little interaction among musicians—the agents in our model system. As interactive behavior emerges among musicians, the joint activity of the interacting ensemble in the Coordinated Performance Events (Introduction and Finale) begin to show increased recurrent points overall (Recurrence Rate), with emerging presence of longer sequences of behavior as represented by higher levels of Determinism, and higher levels of Entropy indicating more variability in sequence length. Increased values of Laminarity indicate enhanced stability in the system. The intermittency of these stable periods in the Coordinated Performance Events as shown in the recurrence plots can also be seen in the varying high and low values of Laminarity over time.

Statistical Analysis

Descriptive Statistics

Statistical analysis was performed on the first and last 25 samples of each Coordination Category, representing the first and last 30 seconds each of Uncoordinated and Coordinated sections of the piece. Recurrence (REC) for the first thirty seconds of each category was highest in coordinated (mean 0.093/ sd 0.030) when compared to uncoordinated (mean 0.081/ sd 0.009). REC for Performance Event showed a slight decrease from Aleatoric (mean 0.085 /sd 0.009) to Transition (mean 0.078 / sd 0.008) and increased in both Introduction (mean 0.089/ sd 0.017) and Finale (mean 0.098 /sd 0.039) (Figure 4.3 A). Determinism (DET) for the first thirty seconds of each category was, highest in the coordinated condition (mean 0.277/ sd 0.147) when compared to uncoordinated (mean 0.163/ sd 0.024). DET for Performance event shows the same pattern, with a slight decrease from Aleatoric (mean 0.167 /sd 0.023) to

Transition (mean 0.160 / sd 0.025), and increases in both Introduction (mean 0.266/ sd 0.106) and Finale (mean 0.288 /sd 0.179) (Figure 4.3, B). Entropy for the first thirty seconds of each category was highest in the coordinated condition (mean 0.693/ sd 0.432) when compared to uncoordinated (mean 0.295/ sd 0.091). Entropy for Performance event shows the same pattern, with a slight decrease from Aleatoric (mean 0.305 /sd 0.074) to Transition (mean 0.292 / sd 0.110), and increases in both Introduction (mean 0.654/ sd 0.325) and Finale (mean 0.731 /sd 0.521) (Figure 4.3, C.) Laminarity for the first thirty seconds of each category was highest in the coordinated condition (mean 0.340/ sd 0.175) when compared to uncoordinated (mean 0.236/ sd 0.045). Laminarity for Performance event shows the same pattern, with a slight decrease from Aleatoric (mean 0.240/ sd 0.053) to Transition (mean 0.234 / sd 0.037), and increases in both Introduction (mean 0.353/ sd 0.162) and Finale (mean 0.326 /sd 0.190) (Figure 4.3, D).

Higher values and increased variability were evident for most RQA measures for Performance Events within the Coordinated sections of the performance vs Uncoordinated sections (Figure 4.3). However, distribution plots do not readily visualize different trajectories of behavior over time. Varying trajectories of each RQA measure for Coordinated vs Uncoordinated sections of the performance is evident in the serial plots in (Figure 4.4). The majority of RQA values hover around a single value over time during the Uncoordinated Performance Events (Aleatoric and Transition), indicating little interaction among musicians—the agents in our model system. As interactive behavior emerges among musicians, the joint activity of the interacting ensemble in the Coordinated Performance Events (Introduction and Finale) begin to show increased recurrent points overall (Recurrence Rate), with emerging presence of longer sequences of behavior as represented by higher levels of Determinism, and higher levels of Entropy indicating more variability in sequence length. Increased values of Laminarity indicate enhanced stability in the system. The intermittency of these stable periods in the Coordinated Performance Events as shown in the recurrence plots can also be seen in the varying high and low values of Laminarity over time.

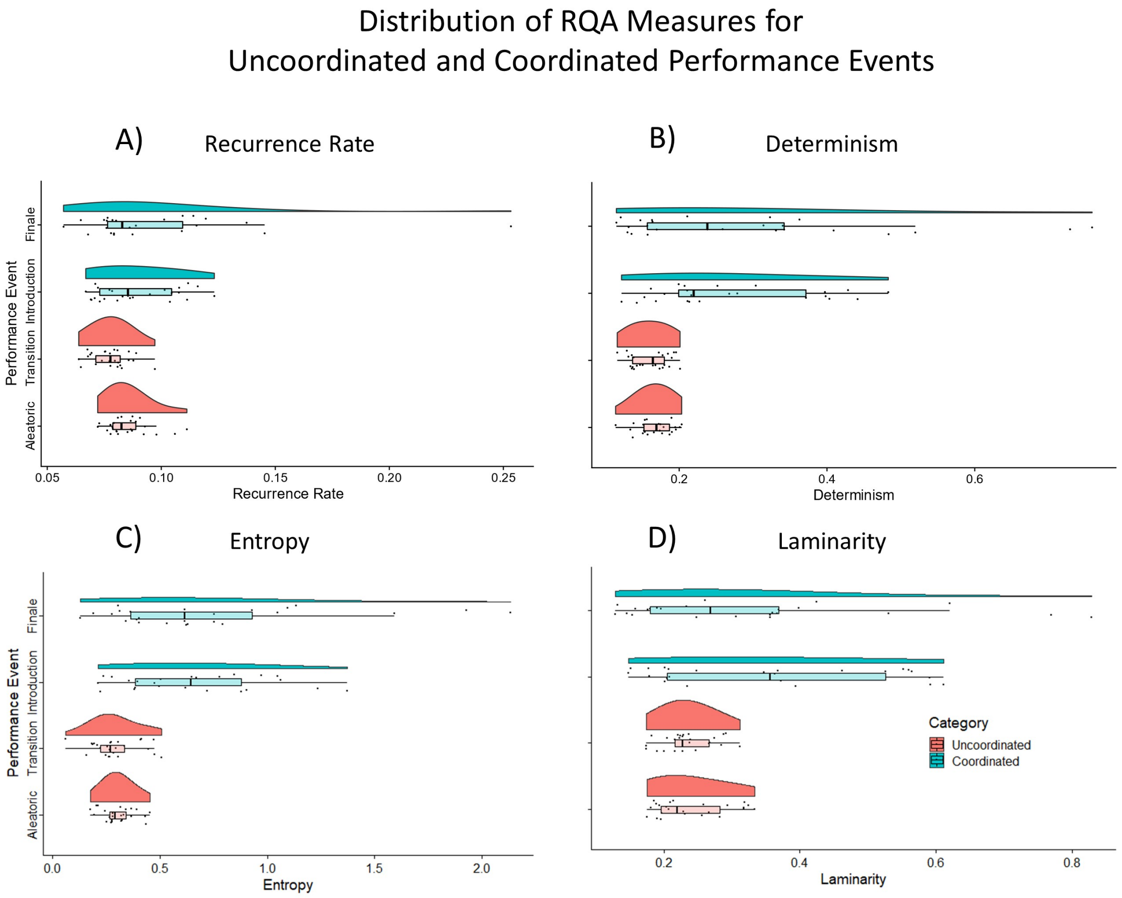


Figure 4.3: Raincloud Plots show higher levels, and wider variance, in each RQA metric in Coordinated compared to Uncoordinated Categories. Boxplots show sample median and interquartile range. Note: 9 seconds of audience applause during the Introduction Performance Event were not analyzed, and are subsequently excluded from all data visualizations.

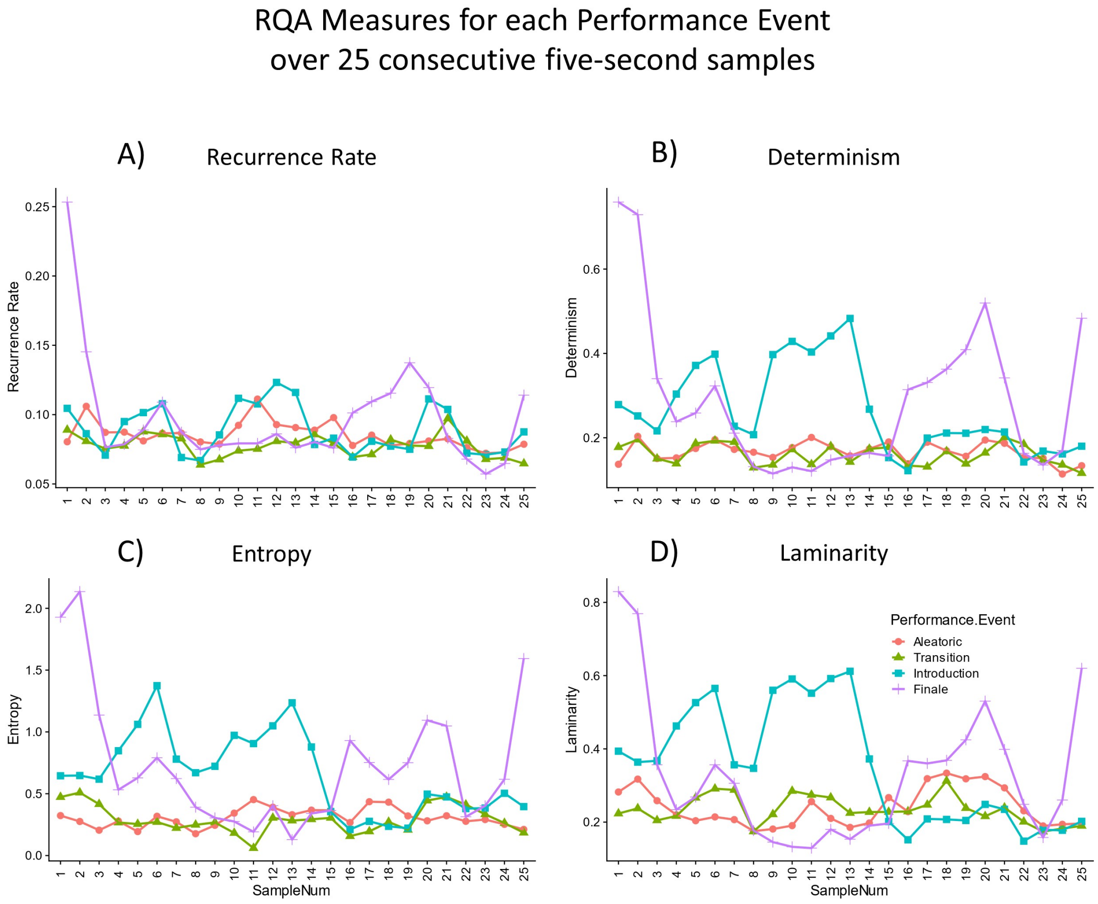


Figure 4.4: Serial plots visualizing the trajectories of recurrence behaviors over time. A) Coordinated Performance Events (Introduction and Finale) show increased recurrent points overall (Recurrence Rate) compared with Uncoordinated Performance Events (Aleatoric and Transition). B) Emerging presence of longer sequences of behavior as represented by higher levels of Determinism, and C) higher levels of Entropy in Coordinated Performance Events, indicating more variability in sequence length. D) Increased values of Laminarity in Coordinated Performance Events indicate enhanced stability in the system. The intermittency of these stable periods in the Coordinated Performance Events as shown in the recurrence plots (Figures 4.1 and 4.2) can also be seen in the varying high and low values of Laminarity over time in the serial plots. Note: 9 seconds of audience applause during the Introduction Performance Event were not analyzed, and are subsequently excluded from all data visualizations.

Inferential Statistics - Model Comparisons

To further examine the trends above, Linear Mixed Effects models (LMEs) were applied to determine the differential effect of Coordination Category and Performance Event on each RQA measure of interest. LMEs (or multilevel models) account for the nested structure of hierarchical data, as when individual observations are nested within groups (Demos, Chaffin, & Logan, 2017). In this case, the individual observations of RQA measures in each 5 second sample are nested within larger uncoordinated or coordinated categories (or the subcategories of performance event). A linear mixed effects model assumes that fixed effects (of coordination category or performance event) do not vary, while the random effect structure of a LME allows each individual sample to vary. This accounts for interdependence between subsequent samples in each category. Model comparisons between LME with and without fixed effects enables inference regarding the contribution of the fixed effect of interest (Winter, 2013). If model comparisons show that a model with fixed effects is statistically different from a model without a fixed effects (a random effects only model in this case), then we can conclude that the model with the fixed effect better explains the data. Therefore, we can infer the differential effect of the fixed effect of interest (Coordination Category or Performance Event) on the RQA measure of interest.

Log REC was predicted by Coordination Category (χ^2 (1) = 8.0805, p = 0 .0045), and by Performance Event (χ^2 (3) = 12.594, p = 0.0056). Log DET was predicted by Coordination Category ($\chi^2(1)$ = 30.455, p = 3.418e-8), and by Performance Event (χ^2 (3) = 30.656, p = 1.004e-6). Log Entropy was predicted by Coordination Category (χ^2 (1) = 41.261, p = 1.332e-10), and by Performance Event (χ^2 (3) = 41.739, p = 4.557e-9). Log Laminarity was predicted by Coordination Category (χ^2 (1) = 10.382, p = .0013), and by Performance Event (χ^2 (3) = 11.516, p = 0.0092).

LME model comparison results are reported in Table 4.1 for log transformations of each RQA measure, except for Average Diagonal Length because assumptions of normality/heterocedasticity were not met. LMEs with a fixed effect of Coordination Category showed lower AIC and BIC values than the null model, or the model with a fixed effect of Performance Event, indicating that RQA measures are better predicted by the Coordination Category (Uncoordinated vs Coordinated) of each sample than by the Performance Event (Aleatoric, Transition, Introduction, and Finale—which are smaller subdivisions of each Coordination Category). Table 4.1. Linear mixed effects model results. Models were fixed effect of interest (Coordination Category or Performance Event) with random effect of Sample Number, against a random intercept model without the fixed effect in question. Models reveal differential effect of the fixed effect of interest on the log transformed RQA measure of interest (Recurrence Rate, Determinism, Entropy, or Laminarity).

Table 4.1: Linear Mixed Effects models evaluating the effect of Coordination Category and Performance Event on each RQA Measure of Interest

Fixed Effects	**Recurrence Rate** Full Model		Null Model	**Determinism** Full Model		Null Model
	Estimate	(SE)		Estimate	(SE)	
Coordination Category						
Uncoordinated (Intercept)	-2.514	(0.0304)		-1.8225	(0.0503)	
Coordinated	0.1052	(0.0360)		0.4217	(0.0703)	
Goodness of Fit						
Deviance	-43.82		-35.74	75.801		106.256
AIC	-35.82		-29.74	83.801		112.256
BIC	-25.4		-21.924	94.222		120.071
χ^2 (df)	8.0805(1)**			30.455(1)***		
Performance Event						
Aleatoric (Intercept)	-2.4678	(0.0389)		-1.8006	(0.0706)	
Transition	-0.0924	(0.049)		-0.0439	(0.0993)	
Introduction	0.0324	(0.0494)		0.4039	(0.0993)	
Finale	0.0856	(0.0494)		0.3956	(0.0993)	
Goodness of Fit						
Deviance	-48.334		-35.74	75.599		106.256
AIC	-36.334		-29.74	87.599		112.256
BIC	-20.703		-21.924	103.23		120.07
χ^2 (df)	12.594(3)**			30.656(3)***		

Fixed Effects	**Entropy** Full Model		Null Model	**Laminarity** Full Model		Null Model
	Estimate	(SE)		Estimate	(SE)	
Coordination Category						
Uncoordinated (Intercept)	-1.2647	(0.0709)		-1.4567	(0.0535)	
Coordinated	0.7169	(0.1003)		0.2506	(0.0758)	
Goodness of Fit						
Deviance	145.78		187.04	89.655		100.037
AIC	153.78		193.04	97.655		106.037
BIC	164.2		200.86	108.08		113.85
χ^2 (df)	41.261(1)***			10.382(1)**		
Performance Event						
Aleatoric (Intercept)	-1.2167	(0.1001)		-1.4504	(0.0753)	
Transition	-0.0961	(0.1415)		-0.0127	(0.1065)	
Introduction	0.6593	(0.1415)		0.3008	(0.1065)	
Finale	0.6783	(0.1415)		0.1877	(0.1065)	
Goodness of Fit						
Deviance	145.3		187.04	88.521		100.037
AIC	157.3		193.04	100.52		106.04
BIC	172.94		200.86	116.15		113.85
χ^2 (df)	41.739(3)***			11.516(3)**		

Note. ***p<0.001, **p<0.01, *p<0.05

4.5 Discussion

The current study evaluated a musical performance which enacted a real-life simulation of the transition from uncoordinated to orchestrated (non-emergent) coordinated behavior. That is, the musicians in this ensemble simulated the transition from disorder to order in one form of social interaction–a musical performance–by following the orchestration of the musical score, accentuated by the presence of a conductor on stage at just the time the musicians begin to play music *together* as a single interacting ensemble.

We have empirically demonstrated differences in the acoustic coordination patterns of this originally non-interacting collection of independent musicians versus their collective dynamics as an interdependent group of musicians participating in a joint musical interaction. Thus, this study is one example of how conceptualizing and evaluating musical interaction using the tools of coordination dynamics and dynamical systems theory can reveal insights into the self-organizing behavior which underlies multi-agent musical interaction (see Schiavio et al, 2021 for a thorough review (Schiavio, Maes, & van der Schyff, 2021)).

Unlike previous studies of human social interaction, this study evaluated acoustic data of a musical performance to infer the global behavior of musicians in the performing ensemble. This means we did not have access to the behavior of individual musicians in order to evaluate correlations between individuals. Instead, we relied on recurrence features extracted from global acoustic data and represented in recurrence plots. The uncoordinated acoustic behavior of individual, non-interacting musicians at the start of the performance demonstrated lower measures of recurrence and stability in the group-level acoustic data. When the musicians began interacting with each other as a coordinated musical ensemble, recurrence and stability measures increased overall. The interaction served as a coupling mechanism for this multi-agent human group, just as the physical platform served as a coupling mechanism for the metronome group. In addition, the interacting ensemble also exhibited further variability in these recurrence measures. This indicates that the interacting ensemble of musicians were able to explore a greater variety of acoustic behavior than when they performed as a stage full of individual musicians.

When the function of each musician becomes interdependent with the functions of the other musicians, the group becomes a complex system. The collective cognition (c.f. (Clark, 2003; Theiner, Allen, & Goldstone, 2010)) that takes place in order to generate the coordinated music makes the group function a little bit like one large mind (Kirchhoff & Kiverstein, 2020; Spivey, 2020) (c.f. (O'Regan & Noë, 2001)). Thus, when that collective mind has an audio time series extracted from it (that is subjected to state-space reconstruction),it can provide insight into the dynamics by which those individual subcomponents achieve their coordinated behavior.

We present an example of Recurrence Quantification Analysis applied to group level acoustic data from a single performance. However, this musical performance is a model system for other forms of multi-agent human interaction. A priori, we know the local rules that govern the emergence of interaction in this ensemble arise from a musical score, which stipulates when the musicians must begin performing as an interacting group, as well as a conductor who acts as a leader during the coordinated section of musical performance. This provides for the ecological validity of a natural performance as well as ground truth knowledge of the ensemble's acoustic performance as they transition from uncoordinated to coordinated behavior. Thus, RQA applied to this model system provides an indication of the possible expected distributions for what recurrence dynamics to expect in truly emergent coordination in multi-agent human interaction in the wild–perhaps in less orchestrated (i.e. more improvised) forms of musical interaction such as leaderless interaction in free jazz improvisation(Goupil, Saint-Germier, Rouvier, Schwarz, & Canonne, 2020), and even day-to-day social dynamics extending beyond musical interaction, such as walking in groups(Tunçgenç, Travers, & Fairhurst, 2021) or interacting in large crowds.

Extending the current analysis methods to other forms of multi-agent human interaction will also expand current knowledge regarding the affective dynamics of acoustic and motor coordination during social interaction. Listening to music while moving in time with a partner increases perceived connectedness among a dyad (Demos, Chaffin, Begosh, Daniels, & Marsh, 2012). Movement synchrony in dancers increases affiliation with the group (von Zimmermann, Vicary, Sperling, Orgs, & Richardson, 2018) and can increase affective engagement from an audience (Vicary, Sperling, von Zimmermann, Richardson, & Orgs, 2017). In a dot-motion paradigm, velocity-based synchrony (associated with expert interaction) in comparison with interval-based synchrony (associated with novice interaction) from ostensibly improvising performers is rated by observers as more beautiful, and the 'performers' are judged to like each other more (McEllin, Knoblich, & Sebanz, 2020). The affiliatory effects of synchronous interaction are not always positive, however, and can actually lead to increased compliance with requests to engage in aggressive behavior (Wiltermuth, 2012). Multi-agent groups in the wild, such as crowds at a sporting event or gathering for a protest, may not be engaging in strictly synchronous motor coordination, however their acoustic behavior may exhibit measurable patterns of distributed coordination which influence the affective states of the group and individual. A question remains as to what are the recurrence dynamics of acoustic behavior of multi-agent groups in the wild, and what role does coordinated acoustic behavior play in the affective dynamics of individuals engaging in or observing these social interactions.

4.6 Methods

Extracting Acoustic Data

Table 4.2: Coordination Category and Performance Event Labels for Recurrence Quantification Analysis. 30-second samples used for analysis are indicated. Audience noise was discarded before analysis.

Music Event	**Coordination Category**	**Performance Event**	**Start Time**	**End Time**	**30-second Sample**
Recording Starts	Audience Noise	–	0m00s	0m14s	–
Scattered Entrances	Uncoordinated	Aleatoric	0m14s	2m43s	0m14s to 0m44s
Flute Cue	Uncoordinated	Transition	2m43s	3m15s	2m45s to 3m15s
French Horn Cue	Coordinated	Introduction	3m15s	3m53s	3m15s to 3m53s*
Conductor Appearance	Audience Noise	–	3m37s	3m45s	**discarded above*
Drum Cue	Coordinated	–	3m53s	9m06	–
Performance Continues	Coordinated	Finale	8m36s	9m06s	8m36s to 9m06s
Performance Ends	Audience Noise	–	9m06s	9m23s	–

An audio recording of ‘Welcome to the Imagination World’ was obtained from the 2009 performance by the Inagauken Wind Orchestra posted on YouTube. An MP3 was downloaded using the YouTube to Mp3 video converter. The audio recording was labeled by two independent raters with terminal music degrees and substantial training in music theory. Raters were familiarized with the program notes for the composition (retrieved from www.windrep.org) and were instructed to identify where the musicians’ transition from “random ad lib” to “unify[ing] majestic introduction” as described in the program note from the composer (operationalized as *uncoordinated* and *coordinated*, respectively), as well as noting any details of the performance they found important. The audio was subsequently labeled into two Coordination Categories: Uncoordinated and Coordinated. The first and last 30-seconds of each Coordination Category (Uncoordinated or Coordinated) were also labeled into four Performance Event subcategories (Aleatoric, Transition, Introduction, and Finale), two in each Coordination Category, respectively, as shown in Table 4.2.

The audio recording was converted from stereo to mono in Audacity 2.3.0, converted from an MP3 to a WAV file, removed DC offset, and normalized to -1.0 dBFS. Python 3.7 in Jupyter Notebooks was used first to create a time series of the full audio data, then to downsample this time series from 44.1KHz to 44.1 Hz. Downsampling to this rate prioritizes the rhythmic content and aggregate amplitude of the acoustic signal rather than pitch or harmonic properties for the purposes of Recurrence Quantification Analysis. It may be a concern that this is a low sampling rate in relation to human auditory perception, which is sensitive to pitch information in the 20hz to 20,000hz range. This is not problematic, however, as this analysis does not seek to explain pitch perception but rather recurrence properties of sound onsets in the acoustic signal itself. This downsampling filters out sound wave properties interpreted as pitch by the human auditory system while preserving frequencies relevant to rhythm perception and identification of event sequences. A 44.1hz sample rate is more than sufficient to capture rhythmic events performed within a tempo range of 60bpm to 135bpm (1hz to 2.25hz) as in this performance. Finally, a separate time series was created for each 30-second Coordination Category. The time series for each labeled Performance Event Category was then extracted into 5-second overlapping windows, sliding by 1 second at a time, saving only full 5-second samples. Time series containing only audience noise, or dominated by audience noise, were discarded. This included nine seconds within the 'Introduction' Coordination Category which were discarded due to noise from audience applause overshadowing the signal from the music. These samples are not included in any analysis or data visualization.

Note: 30 second samples were chosen for analysis due to limitations in duration of uncoordinated performance. The Uncoordinated section of music, at 3 minutes in duration, is half the length of the remaining 6 minutes of the Coordinated section. An equal representation from each Coordination Category is required so as not to bias the results of statistical analysis. We know, due to the score, where the coordinated music begins and ends. Further, independent raters indicated a clear Transition section during the last thirty seconds of Uncoordinated performance. For this reason we choose to analyze both the first and last thirty seconds of the shorter Uncoordinated section, and the first and last thirty seconds of the longer Coordinated section. This selection also allows us to compare recurrence dynamics between the the start and end of each Coordination Category in the case that RQA shows individuals are more coordinated at the end of a section than the beginning after interacting for a period of time, rather than solely because of the coordination indicated by the musical score. For an overview of the global variability of recurrence metrics across a larger subset of data, see Supplementary Figure 1 for a serial plot visualizing all six 30-second samples from Uncoordinated performance and the first and last three 30-seconds of Coordinated performance.

Recurrence Quantification Analsys

The CRP toolbox in MATLAB 2018b was used to visualize the acoustic time series as recurrence plots and to carry out RQA (Marwan, n.d.; Marwan, Romano, Thiel, & Kurths, 2007). RQA parameters were set with an embedding dimension of 4, delay of 10, neighborhood size (radius) of 1 $*$ standard deviation, using maximum norm to calculate neighbors of the phase space trajectory. Parameters for the time delay and embedding dimension were chosen based on AMI and FNN respectively using a custom MATLAB GUI provided from the 2019 APA Advanced Training Institute in Nonlinear Methods for Psychological Science. There are various approaches to setting the threshold value for detecting nearest neighbors. In classification based on recurrence dynamics of harmonic, transient, and noisy acoustic signals, Zhang 2011 (Zhang, Liu, Zhang, & Bu, 2011) set this threshold value using 1 $*$ standard error. Here we set the threshold value at 1 $*$ standard deviation, because the standard deviation is always larger than standard error, assuring a radius large enough to sufficiently capture recurrent structures in the recurrence plots. $5 * \sigma$ has been suggested as an optimal threshold value for detecting signal in cases of high observational noise (Thiel et al., 2002), however, $1 * \sigma$ is standard and is preferable when the amount of observational noise is unknown (Schinkel, Dimigen, & Marwan, 2008). For further discussion regarding parameter selection in RQA see Marwan, 2011 (Marwan, 2011) and Webber and Marwan, 2015 (Webber & Marwan, 2015).

Statistical Analysis

Statistical analysis was performed on the first and last 25 samples of each Coordination Category, representing the first and last 30 seconds each of Uncoordinated and Coordinated sections of the piece. Raincloud plots (Allen et al., 2021) and serial time series plots to visualize distributions and trajectories of the RQA measures were created in RStudio 1.1.463 using ggplot2 (Wickham, 2016).

Linear Mixed Effects models (LMEs) were applied to determine the differential effect of Coordination Category and Performance Event on each RQA measure of interest. LMEs were calculated using the lme4 package (Bates, 2015). The first model examined the effects of Coordination Category on each RQA measure, with a fixed effect of Coordination Category and random effects of Sample Number, to account for any variance arising from individual five-second samples. The second model examined the effects of Performance Event on each RQA measure, with a fixed effect of Performance Event and random effects of Sample Number.

- Full model with fixed effect of Category (Coordinated vs Uncoordinated) and random effects of order (sample number):

$$\begin{aligned} \text{RQAMetric}_i &\sim N\left(\alpha_{j[i]} + \beta_1(\text{Category}), \sigma^2\right) \\ \alpha_j &\sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{ for SampleNum j} = 1, \ldots, \text{J} \end{aligned}$$

- Null, intercept-only model without the fixed effect of Category:

$$\begin{aligned}\text{RQAMetric}_i &\sim N\left(\alpha_{j[i]}, \sigma^2\right)\\ \alpha_j &\sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{ for SampleNum j} = 1, \ldots, \text{J}\end{aligned}$$

- Full model with fixed effect of Performance Event (Aleatoric, Transition, Introduction, Finale) and random effects of order:

$$\begin{aligned}\text{RQAMetric}_i &\sim N\left(\alpha_{j[i]} + \beta_1(\text{Performance. Event}), \sigma^2\right)\\ \alpha_j &\sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{ for SampleNum j} = 1, \ldots, \text{J}\end{aligned}$$

- Null, intercept-only model without the fixed effect of Performance Event:

$$\begin{aligned}\text{RQAMetric}_i &\sim N\left(\alpha_{j[i]}, \sigma^2\right)\\ \alpha_j &\sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{ for SampleNum j} = 1, \ldots, \text{J}\end{aligned}$$

LMEs with random intercepts such as this are robust to variability in individual subjects, or five-second samples in this case. This is because random intercept models assume a different baseline-level of the RQA measure of interest in each fixed effect for each sample, thus accounting for any differences that may appear by virtue of the sequential order of obtaining each sample. Mixed models also address issues of non-independence due to inherent correlations between successive samples of musical performance data (Demos et al., 2017). Goodness of fit was evaluated by model comparison of the full models against null, intercept only models without the fixed effect in question, as shown in the equations above. Four of the five RQA measures of interest were modeled Recurrence Rate, Determinism, Entropy, and Laminarity). LME results based on Average Diagonal Length are not reported. Residuals plots revealed that the LMEs for Average Diagonal Length did not meet criteria for assumptions of normality and heteroscedasticity, even after log transformation, and as such were not a good model for the data. Statistical significance was obtained by computing a likelihood ratio test of the full model to a null model without the fixed effect in question.

Acknowledgements

The authors wish to thank Byron Pillow, M.M. (National Music Museum, University of South Dakota) for providing technical assistance and music theory consultation, and Dr Paul Garza, D.M.A. (University of Houston) for music theory consultation.

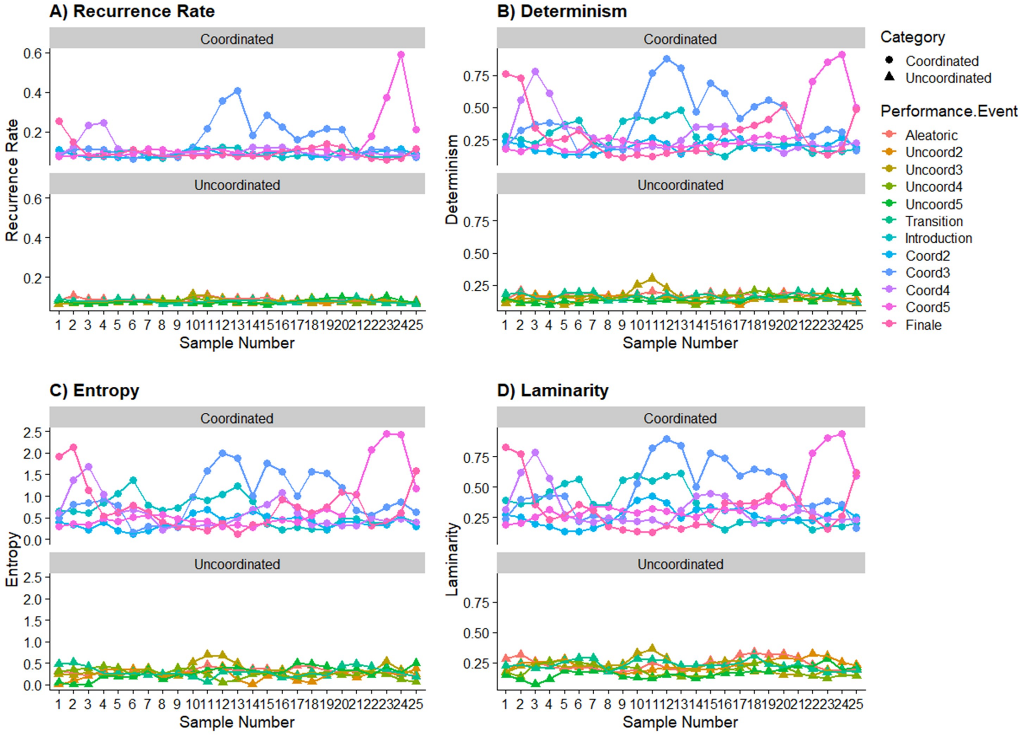


Figure 4.5: Supplemental Figure. Serial plots visualizing the trajectories of recurrence behaviors over time. Shown are all six 30-samples of Uncoordinated performance, and the first and last three 30-second samples of Coordinated performance. Recurrence and stability measures are increased overall and exhibit greater variability during Coordinated compared to Uncoordinated performance.

This material is based upon work supported by the National Science Foundation under Grant No. DGE-1633722. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author contributions statement

S.P. devised the project, the main conceptual ideas, and led writing the manuscript. M.R. performed data cleaning and preparation. S.P. and M.R. performed the nonlinear analysis, S.P performed statistical analysis. R.B. and M.S. verified the analytical methods and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Chapter 5

Recurrence Quantification Analysis of Crowd Sound Dynamics at a Basketball Game

Chapter 5 is an unpublished empirical study, investigating the acoustical behavioral dynamics of crowd sound at a basketball game. This chapter extends the theories and methods presented in chapters three and four, and applies the same nonlinear statistical analysis techniques to investigate the interaction of fans over the course of a basketball game. When applying these analysis methods to the acoustical signal generated by the crowd, we can identify differing coordination patterns in different forms of crowd behavior, such as cheering, chanting, or making sounds to distract the other team. We can further use the results from these analyses to train a model that classifies between these forms of crowd sound. The results of this study provide further support for the use of these techniques in analyzing group-level behavior. The chapter concludes by contextualizing these results within the larger research program of joint action and coordination, and proposes future studies extending to interaction over multiple games and at different sporting events.

Authors:

Shannon Proksch, Majerle Reeves, Kent Gee, Mark Transtrum, Chris Kello, and Ramesh Balasubramaniam, *unpublished*

Abstract

When multiple individuals interact in a conversation or as part of a large crowd, emergent structures and dynamics arise that are behavioral properties of the interacting group rather than of any individual member of that group (M. A. Riley et al., 2011). Recent work using traditional signal processing techniques and machine learning has demonstrated that global acoustic data recorded from a crowd at a basketball game can be used to classify emergent crowd behavior in terms of the crowd's purported emotional state (Butler et al., 2018). We propose that the description of crowd behavior from such global acoustic data could benefit from nonlinear analysis methods derived from dynamical systems theory. Following from Zhang et al., 2011 and Proksch et al., 2022, we leverage nonlinear analyses to extract features that are relevant to the behavioral interactions that underlie audio signals produced by a crowd. We propose that recurrence dynamics measured from these audio signals via Recurrence Quantification Analysis (RQA) reflect information about the behavioral dynamics of the crowd itself. We analyze these dynamics from acoustic signals recorded from a crowd at a basketball game, and that were manually labeled according to the crowd's emotional state across six categories: angry noise, applause, cheer, distraction noise, positive chant, and negative chant. We show that RQA measures are useful to differentiate the emergent acoustic behavioral dynamics between these categories, and can provide insight into the recurrence patterns that underlie crowd interactions.

Keywords: Recurrence Quantification Analysis; Acoustical Analysis; Crowd Behavior; Dynamical Systems; Emergence

5.1 Background

Interaction-dominant systems – such as a collection of individuals interacting within a crowd – can be described by emergent structures and dynamics that are behavioral properties of the system itself, rather than of any individual component (M. A. Riley et al., 2011). The emergent dynamics of crowd behavior have been fruitfully modeled according to biological phenomena such as swarm behavior (Kok, Lim, & Chan, 2016). Classification of emergent crowd dynamics, often using computer vision technology, have typically relied on analysis of video data for features such as crowd density estimation, motion detection, and movement/behavior tracking of individual signals or group behavior (Kok et al., 2016; Swathi, Shivakumar, & Mohana, 2017). However, it is not always feasible to obtain high quality image, video, or speech data of a crowd in action, nor is it always feasible to obtain signals measured from each individual in an interacting crowd. We seek to extend the study of emergent crowd behavior to include analysis of the global acoustic output of a crowd as a whole. This acoustical analysis of crowd behavior can augment current video-based crowd behavior analysis, and can also mediate in cases where video data are incomplete or unclear.

Recent work using traditional signal processing techniques and machine learning has demonstrated that global acoustic data recorded from a crowd at a basketball game can be used to classify crowd behavior in terms of the crowd's purported emotional state (Butler et al., 2018). Importantly, these data were not a collection of individual acoustic signals from individual members of the crowd, but rather a global acoustic signal measured from the crowd as a whole. Common acoustic analyses used for classifying human speech, as well as crowd noise, include spectral and mel frequency cepstral coefficients (Reddy, Sinha, & Seshadri, 2013), where the latter measure is based on an approximatation of human hearing (Singh & Rani, 2014). These measures assume that at short enough time scales important features of an audio signal are reasonably stationary. Nonlinear analysis techniques, such as Recurrence Quantification Analysis (RQA), are adept at capturing exactly that nonstationarity that characterizes audio signals at longer time scales (Wallot & Leonardi, 2018). We propose that analysis of crowd behavior from such global acoustic data could benefit from taking a dynamical systems approach that embraces the nonlinearity and nonstationarity present in the sounds generated by an interacting crowd.

5.2 Dynamical Systems

Dynamical systems theory seeks to describe the nonlinear behavior of large scale systems that emerges from interacting components/individuals (Connell, DiMercurio, & Corbetta, 2017). Such emergent behavior arises due to the soft-assembly of individual components into metastable patterns of behavior (Kello & Van Orden, 2009). When large groups of people gather together, they consciously and unconsciously coordinate their behavior in a number of ways, from cheering with the same chants to spontaneously synchronizing in their applause at concert. The patterns of synchronicity in the sounds generated by crowds demonstrate a process of social self-organization (Néda et al., 2000).

One tool in the dynamical systems toolbox is Recurrence Quantification Analysis (RQA). RQA is used to quantify structures that can be visualized in recurrence plots generated from the nonlinear behavior of a time series that has been subject to state space reconstruction (Marwan, Wessel, Meyerfeld, Schirdewan, & Kurths, 2002; Marwan et al., 2007; Takens, 1981; Vlachos & Kugiumtzis, 2010). Traditional RQA, as well multivariate approaches like crossRQA and multidimensional RQA (mdRQA), has proven useful in describing behavioral aspects of joint action in dyadic and group interaction. These analyses are robust to the nonlinearity and nonstationarity of time-dependent signals, and can be used to evaluate relative coordination dynamics as well as transitions between order and chaos in such systems (c.f. Wallot and Leonardi, 2018 for a detailed review and tutorial).

In a recurrence plot, time series data are plotted on axes of time by time. A point $(i,\ j)$ is plotted if the value at time i and time j are recurrent within a specified neighborhood size of an N-dimensional state-space after state space reconstruction. The line of incidence (LOI) along the main diagonal shows the time series at a time lag of 0. Each step away from the LOI represents the trajectory of the system at a time lag, depicting self-similarity of the system over time.

Information from a recurrence plot is quantified into a variety of metrics via RQA. The recurrence rate quantifies the percentage of points on a recurrence plot and represents patterns of behavior that persist over time. Determinism quantifies the percentange of points that fall on any diagonal line in the plot (except the LOI), and represents behaviors that belong to a particular pattern of behavior over time. Entropy is the variability in these line lengths, representing disorder of these sequences. Finally, laminarity quantifies the percentage of points that fall on a vertical line on a recurrence plot, and represents clusters of behavior over a short period of time to which the system may temporarily visit, leave, and return. Examples of determinism and laminarity depicted from two samples of basketball crowd sound data are shown in Figures 5.1 and 5.2.

Recurrence Quantification Analysis has additionally proven useful in describing and classifying acoustic data. Zhang et al., 2011 made use of recurrence plots and recurrence quantification analysis to classify audio signals into noise-like, transient, harmonic-like, and mixed signals(Zhang et al., 2011). Proksch et al., 2022 further justified the use of RQA to describe differences in acoustical signals generated by multi-agent behavior of a performing musical ensemble who were either uncoordinated with each other (the audio signal they collectively generated was noise-like), or were coordinated with each other (the audio signal they collectively generated was harmonic-like) (Proksch, Reeves, Spivey, & Balasubramaniam, 2022).

5.3 Dynamics of Collective Interaction

The context of a social event affects the emergence of collective synergies, synchronicities, and multistable dynamics in the acoustical behavior of col-

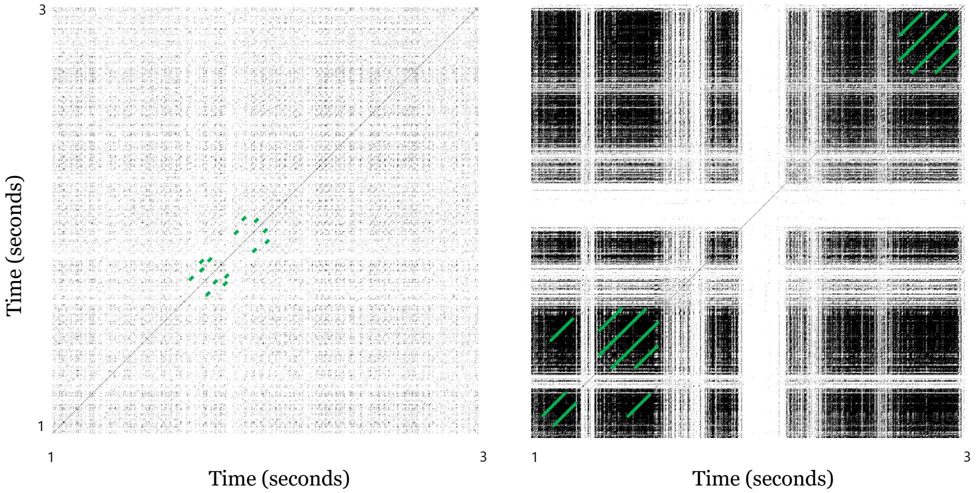


Figure 5.1: Green lines depict trajectories of consecutive recurrent points over time. Longer trajectories are quantified in higher values of determinism. The plot on the left has many short diagonal trajectories consisting of only a few consecutive recurrent points, while the plot on the right contains many long diagonal trajectories consisting of many consecutive recurrent points over time (even more than we have highlighted). These plots are zoomed in on two seconds of data from two different five second samples of crowd sound.

lective interactions. To understand how analysis of the acoustical behavior of a crowd might expand research on joint action and crowd dynamics, it is necessary to provide context on the growing body of research evaluating coordination dynamics that arise during group interactions, much of which has focused on the behavioral and physiological modalities.

Modalities of Synchrony and Coordination in Groups

Recently, Multidimensional Recurrence Quantification Analysis (MdRQA) was used to measure physiological synchrony in the heart rates of fans who attended a live basketball game, and fans who gathered in small groups to attend a live screen of a basketball game on television (Baranowski-Pinto, Profeta, Newson, Whitehouse, & Xygalatas, 2022). MdRQA is able to compute recurrence measures across multiple signals (i.e. the heart rates of multiple individuals), in contrast to RQA that evaluates recurrence measures across the length of a single signal. Increased interdependence in heart rate was found for fans who attended the live game, indicating that there is an enhanced so-

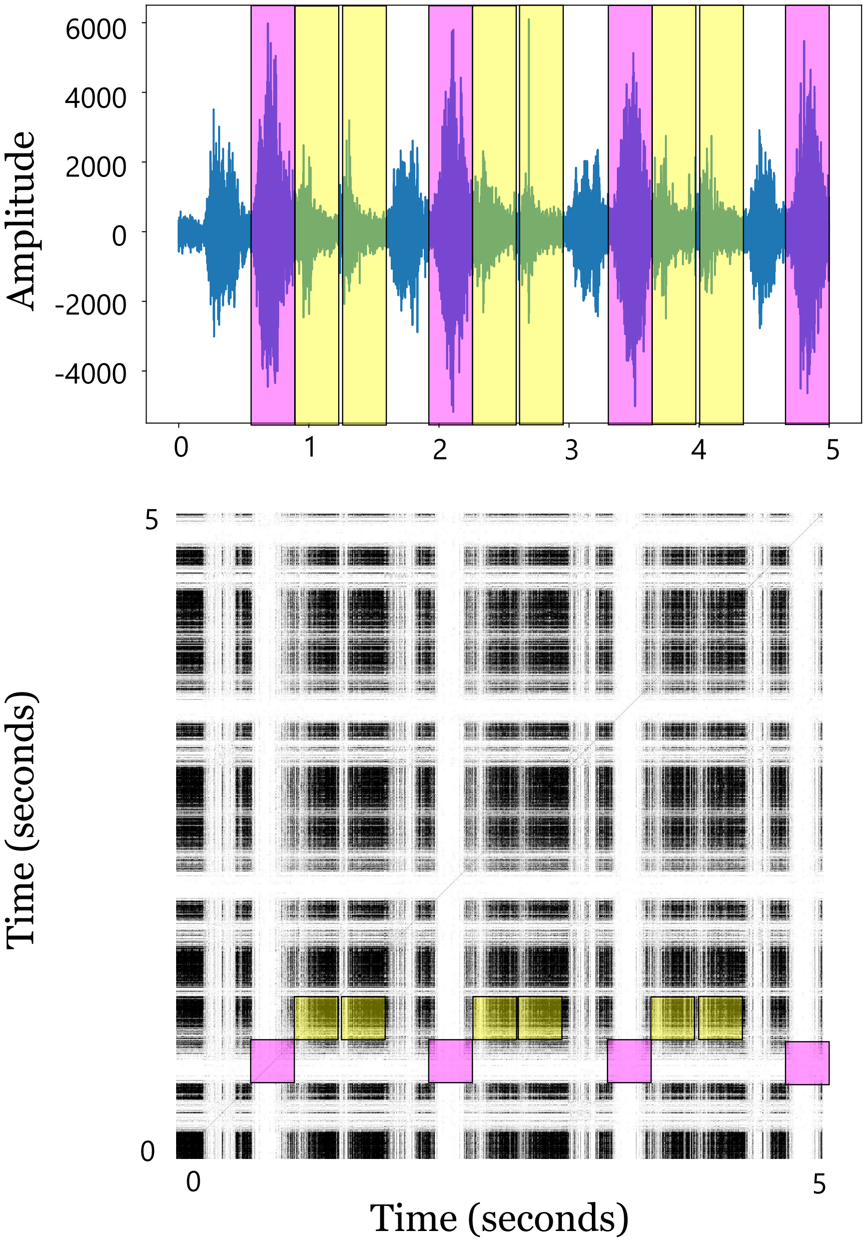


Figure 5.2: Highlighted instances of laminar states in a time series reflected in a recurrence plot. The bursts in the time series appear as white space with few recurrent points in the recurrence plot. This is because this bursty state is revisited only three additional times after the instance highlighted in pink along the main diagonal. Meanwhile, the state highlighted in yellow repeats an additional five times for the duration of the behavior. This plot is one five second sample of crowd sound.

cial effect of the interpersonal dynamics inherent in attending this sporting event live and in-person compared to virtually engaging with the game over a television screen. This was demonstrated through increases in both determinism (DET: indicating stability in the system and the ability to predict future states from past states) and average diagonal line length (ADL: indicating the length of time, or persistence, of recurrent states within the system). Further, Baranowski-Pinto et al., found that individuals who attended the live game exhibited stronger social cohesion, as reported by stronger feelings of transformativeness, or the sense that their individual identity has 'fused' to the identity of the group. These self-report measures were correlated with recurrence measures for fans who attended the game in person. A second study evaluated behavioral synchrony of an audience attending a live or pre-recorded rock concert (Swarbrick et al., 2019). Where basketball fans in Baranoiwski-Pinto et al.,'s study were both watching a live game, which differed only by being in-person or screened on television, Swarbrick et al.,'s study maintained the interpersonal dynamic between concert attendees by having concert-goers in each condition be physically present in the concert venue (Swarbrick et al., 2019). In the live concert, the rock band performed on the stage, while in the non-live condition, a recording of that performance was projected onto the stage. Faster head movements, a measure of vigor and engagement, were found during the live performance than the non-live performance. No effect was found between performances for entrainment with the music – however it was not analyzed whether there was enhanced movement synchrony *between* audience members during either performance.

In a third study, both physiological and behavioral synchrony were evaluated in groups of three people engaged in a joint drumming task (I. Gordon, Gilboa, Cohen, & Kleinfeld, 2020). Synchronous or asynchronous drumming interaction was achieved by asking participants to drum along to an auditory beat with a predictable or unpredictable tempo, respectively. Gordon et al, 2020 found that the drumming task itself led to increased synchrony of heart-beat inter-beat-intervals (IBIs) between group members compared to baseline (non-interaction). Groups with higher physiological synchrony during the initial drumming task were more coordinated during a subsequent free-improvisation drumming task. Interestingly, this increase in heartbeat IBI synchrony was not related to whether the initial drumming task was synchronous or asynchronous, indicating that the enhanced heartbeat IBI synchrony may stem from the effect of the interpersonal interaction itself, rather than the behavioral synchronization of the drumming itself. This may be similar to interaction at a sporting event, where individuals are not always synchronizing directly with other fans in attendance. Further, even during a coordinated cheer individuals may be 'in sync' with a global signal without being ‘in sync’ to other individuals directly nearby.

These studies highlight the importance of interpersonal interaction – as well as a live, in-person interaction context – in facilitating physiological and behavioral synchrony and the emergence and maintenance of shared social bonds. Although the specific modality of focus differed in each study (from physiological measures of heart-rate, to behavioral measures of movement), all of the interactions in these studies occurred in a shared acoustic and auditory environment. Anthropologist and ethnomusicologist, Blake and Cross (2015), state that it is precisely this environment that is ‘one of the most powerful and flexible tools that humans use to manage and mediate relationships with each other and with the environments that they construct or modify’. In the next section, we situate our interest in acoustic social coordination and in the shared social scripts that underlie the coordination of acoustical behavior during the interactions of a crowd.

Acoustic Social Coordination

The emergence of different joint action dynamics can be analyzed in terms of social scripts – implicit or explicit norms for organizing behavior in social contexts that are ‘underwritten by culturally specific narratives’ (Albarracin, Constant, Friston, & Ramstead, 2021). The acoustical behavior of the musical ensemble described in section 1.1 was carried out according to an *explicitly* social script – a musical score– governing the transition from uncoordinated action of individuals to coordinated interaction of a multi-agent group. These two coordination modes are reflected in the emergence of structured recurrence patterns over time (Proksch et al., 2022). Applause after concerts can also follow certain *implicit* social scripts – spreading by initial social contagion (Mann, Faria, Sumpter, & Krause, 2013) and perhaps persisting while slowly dying down, or ending abruptly as soon as a loudness threshold is passed (Michard & Bouchaud, 2005). Fluctuations in the relative synchrony of sound generated by the applause of a crowd attending a classical music concert have been shown to display an emergent periodic signal. Initial applause is fast and asynchronous, and as synchrony increases the overall signal of the sound behavior increases, while the average noise of the sound decreases. This decrease in average noise is a result of a slower clapping period that emerges as individuals clap in unison (Néda et al., 2000). Such audiences fluctuate between asynchronous and synchronous behavior before ultimately fading out as the event draws to a close.

Basketball games are another social context that affords the emergence of coordinated acoustical behavior among a group of interacting people. Rather than an explicitly written script, fans at sporting events follow an at times *explicit* or *implicit* social script, where events in the game and prompts from the announcer or cheerleaders, or other fans, govern the behavior of fans gathered in the arena. Patterns of social self-organization emerge and dissipate according to local interaction among fans and global interactions associated with the game. When your team scores, the social script affords a cheer, when the other team is attempting a free throw, the social script affords generating raucous noises in attempt to distract the player on the court, and when the cheer leaders or a group of fans begin a rehearsed chant ("De-fense"; "B-Y-U Cougars") the social script requires that you chant along. These rehearsed chants are an example of joint speech, a collective phenomenon where multiple individuals repeat the same words simultaneously with the purpose of engaging in group expression (Cummins, 2013).

Synchrony demonstrated in this acoustical behavior, including the synchrony of joint speaking during group chants, is an important characteristic of interpersonal interaction.The repetition of chants or short rhythmic utterances in sporting events enables "synchronized activity...an extreme form of coordination", whereby individuals enact a collective "we" and establish a coordinated group identity for as long as the behavior persists (Cummins, 2020). Joint action research has demonstrated that higher levels of synchronous behavior are associated with various aspects of prosocial cognition including: increased affiliation (Hove & Risen, 2009; Wiltermuth & Heath, 2009), social cohesion (Marsh, Richardson, & Schmidt, 2009), group identity (McNeill, 2022), and cooperation (Kirschner & Tomasello, 2010). A recent meta-analysis has shown that the effects of synchrony on prosocial behaviors and positive affect are larger for larger groups (Mogan, Fischer, & Bulbulia, 2017). However, it can be difficult to measure synchronous activity or joint speech from very large groups of people engaged in naturalistic social interactions. It may not always be possible to obtain one signal from each member of a large group to evaluate correlations and synchrony between those signals. What may be more feasible in such situations, is to record a global acoustic signal generated by the group as a whole.

5.4 Project Aims

Previously, Proksch et al., 2022 applied RQA analysis to a global acoustic recording of a performing musical ensemble. Whether individuals within that ensemble were coordinating their behavior, or not, was dictated by a musical score. These two patterns of behavioral dynamics (uncoordinated vs coordinated) were reflected in RQA metrics derived from analyzing recurrence plots generated from the downsampled audio signals. Here, we analyze a recorded audio signal of a crowd of students engaging in various forms of acoustical behavior at a BYU basketball game (Butler et al., 2018). Following from both Zhang et al., 2011 and Proksch et al., 2022, we leverage nonlinear analyses to extract features that are relevant to the behavioral interaction and coordination of the crowd who produced these audio signals. We propose that recurrence dynamics measured from this global audio signal reflect information about the behavioral dynamics of the crowd itself. We calculate recurrence features using RQA to evaluate the emergent acoustical behavioral dynamics of the interacting crowd.

This paper has two objectives. The first objective is a theory-driven description of crowd sound dynamics using specific RQA metrics relevant for describing system level collective behavior: Recurrence Rate, Determinism, Entropy, and Laminarity (described in more detail in the methods below). We predict that the coordinated acoustical behavior of the crowd will exhibit higher stability and recurrence (measured by determinism and recurrence rate) during pseudo-rhythmically organized joint speech such as rhythmic chants. Meanwhile, the less structured nature of acoustic events such distraction noise will exhibit lower measures of stability and recurrence.

The second objective is a data-driven, machine learning approach to classify each crowd sound based on the full suite of metrics available from the the PyRQA package (a total of 19 RQA metrics, listed in the methods). A combination of RQA metrics and SVM (support vector machine) classification has proven effective at discriminating between nonlinear (and non-stationary) dynamical systems that exhibit similar dynamics. dos Santos et al., 2014 showed that RQA plus an SVM classifier showed successful classification of timeseries data generated from the Logistic map – a canonical example of a nonlinear dynamical system – as well classification of real biological data describing the (nonlinear and non-stationary) dynamics of human heart rate variability across different age groups and health contexts (dos Santos, Barroso, Godoy, Macau, & Freitas, 2014). We apply an SVM classifier on RQA metrics of samples of crowd data that were labeled according to differing classes of acoustical behavior.

5.5 Methods

Crowd Sound Dataset

Our dataset is recorded audio signal from the student section of a single Men's BYU basketball game (Butler et al., 2018). The data set was labeled by BYU undergraduates into different classes of acoustical behavior along with labels of associated game events, shown in Table 5.1.

Table 5.1: Crowd Sound Categories and descriptions from Butler et al., 2018

Crowd Sound Category	Description
Angry Noise	Crowd shouting in anger.
Applause	Crowd clapping that can include crowd vocalization
Cheer	Loud, positive crowd vocalization.
Distraction Noise	Attempts by crowd to draw an opposing team member's attention away from the game, most commonly when the opposing team possesses the ball or is about to shoot a free throw.
Negative Chant	Crowd shouting in anger or distress, usually directed toward referees after a less than ideal call or towards a player from the opposite team.
Positive Chant	Rhythmic crowd shouting, usually directed towards the home team, e.g. "De-fense" or "B-Y-U- Cougars"

We analyze six classes of acoustical behavior that were observed in the crowd during this basketball game: Angry Noise, Applause, Cheer, Distraction Noise, Negative Chant, and Positive Chant. We did not analyze sound events labeled as Singing (which was accompanied by the band or PA system) or Silence. The raw acoustic data were sampled at 50kHz, and we resampled by a factor of ten for nonlinear analysis at 5kHz. As described in Proksch et al. 2022, downsampling by this rate focuses on the higher order rhythmic properties and aggregate amplitude of the acoustic signal, essentially filtering out pitch-information from the acoustical behavior of the audience as well as semi-pitched signals from shoes across the basketball court that may have been picked up by the microphones.

Part 1: Recurrence Quantification Analysis

For Part 1, five-second samples were created using non-overlapping windows (such that a 12 second event will have two 5-second samples, e.g. 0-5s, and 4-9s, and the remaining two seconds in the event are discarded). Any samples shorter than five seconds were discarded, and residual data longer than multiples of five seconds were also discarded. The number of samples for each crowd sound category is reported in Table 5.2:

Table 5.2: Non-overlapping 5s samples used for RQA and linear regression

Crowd Sound Category	**# of samples**
Angry Noise	16
Applause	15
Cheer	47
Distraction Noise	116 (largest class)
Negative Chant	14 (smallest class)
Positive Chant	63

In order to compute statistics comparing balanced classes, data were bootstrapped (oversampled with replacement) so that there are 116 observations in each class.

Based on Proksch et al., 2022, we chose four RQA metrics to evaluate, which are each indicative of different aspects of behavior in nonlinear dynamical systems: Recurrence Rate, Determinism, Entropy, and Laminarity.

- Recurrence rate (the percentage of recurrence points on the recurrence plot) represents patterns of behavior that repeat over time.
- Determinism (the percentage of points that fall on any diagonal line in the recurrence plot) represent behaviors that belong to a longer sequence of behavior
- Entropy (the variability in lengths of these diagonal lines) represents the amount of disorder there is in these sequences
- Laminarity (the percentage of points that fall on a vertical line in the recurrence plot) represents clusters of behavior for a length of time, i.e.

> when a system visits a behavior for a period of time, leaves, and then returns to that behavioral state

RQA was run on the time series data extracted from the downsampled audio of these five-second samples using PyRQA version 8.0.0 (Rawald, Sips, & Marwan, 2017), with an embedding dimension of 5, a delay of 5, and a neighborhood value fixed radius of 1*standard deviation (SD), using maximum norm to calculate neighbors of the phase space trajectory. Parameters for the time delay and embedding dimension were chosen based on AMI and FNN respectively using a custom MATLAB GUI provided from the 2019 APA Advanced Training Institute in Nonlinear Methods for Psychological Science. There are a variety of approaches to setting the neighborhood threshold value. A standard approach, which we use, is setting this threshold value according to a fixed amount of nearest neighbors at some ratio of the standard deviation of the data. This holds constant the number of neighbors within a neighborhood and also hold the number of recurrence points constant in a column of the recurrence plot (Eckmann, Kamphorst, & Ruelle, 1987). It has been suggested that 5*SD be used to accurately detect a signal in the presence of significant observational noise (Thiel et al., 2002). However,"this approach fails for signals of very low signal to noise ratio, or when the amount of noise is unknown" (Schinkel et al., 2008). Additionally, the so-called "noise' is the signal in our data, therefore we settle on a similar approach to Zhang et al 2011 (Zhang et al., 2011), setting this threshold based on a Fixed Amount of Nearest neighbors values of 1*SD. (note: Zhang et al., 2011 (Zhang et al., 2011) used 1*standard error – we chose standard deviation because the value given by SD of the mean is always larger than the SE of the mean, assuring that our radius is large enough to sufficiently capture the recurrence structures in the recurrence plots).

Statistical Analyses

A separate linear regression model with planned contrasts was fit for each RQA metric of interest (Recurrence Rate, Determinism, Entropy, and Laminarity) as a function of six levels of Crowd Sound Category(Angry Noise, Applause, Cheer, Distraction Noise, Negative Chant, and Positive Chant): RQA Metric CrowdSoundCategory. Orthogonal (sum-to-zero) contrasts were used to specify a specific linear combination for each predictor in *a priori* planned comparisons. Since we are not testing a specific treatment or change from any specific baseline of crowd sound behavior, sum contrasts provide the advantage of comparing RQA metrics in each category to average metrics across all categories, rather than to a baseline or control category(Schad, Vasishth, Hohenstein, & Kliegl, 2020). The linear regression and pairwise comparisons of estimated marginal means were implemented separately for each of the four RQA measures of interest. This is because each RQA measure addresses a different aspect of the crowd's behavior over time, as described previously.

Results

RQA Results

Figure 5.3 shows recurrence plots generated from the time series data of a representative five second audio sample from two crowd sound categories. These recurrence plots visualize characteristic patterns of recurrence that are quantified through RQA. Qualitatively, the recurrence plot generated from 5 seconds of distraction noise resembles recurrence plots of uncoordinated group behavior (Proksch et al., 2022) or noisy-like audio signals (Zhang et al., 2011), while the recurrence plot generated from five seconds of labeled positive chant resemble a recurrence plot generated from coordinated group behavior (Proksch et al., 2022) or mixed audio signal (Zhang et al., 2011). Distraction noise shows low levels of stability and recurrence, while positive chant shows high levels of stability and recurrence, as quantified by RQA. Further statistical analysis on the distribution of RQA metrics in each crowd sound category is described below.

Linear Regression with sum-to-zero contrasts and Pairwise Comparisons

Table 5.3 lists the associated descriptive means, median, and standard deviation of RQA metrics, and Figure 5.4 shows the smoothed density distributions and quartile lines for the non-bootstrapped data from each crowd sound category. Overall, the two chant categories (Positive and Negative chant) display the highest values of recurrence, determinism, entropy, and laminarity. The two noise categories (Angry and Distraction Noise) display consistently low values of these RQA metrics, with Angry Noise having particularly low values of Entropy and Laminarity. Cheer and Applause are the most variable, with multimodal or nearly flat distributions exhibited by Cheer.

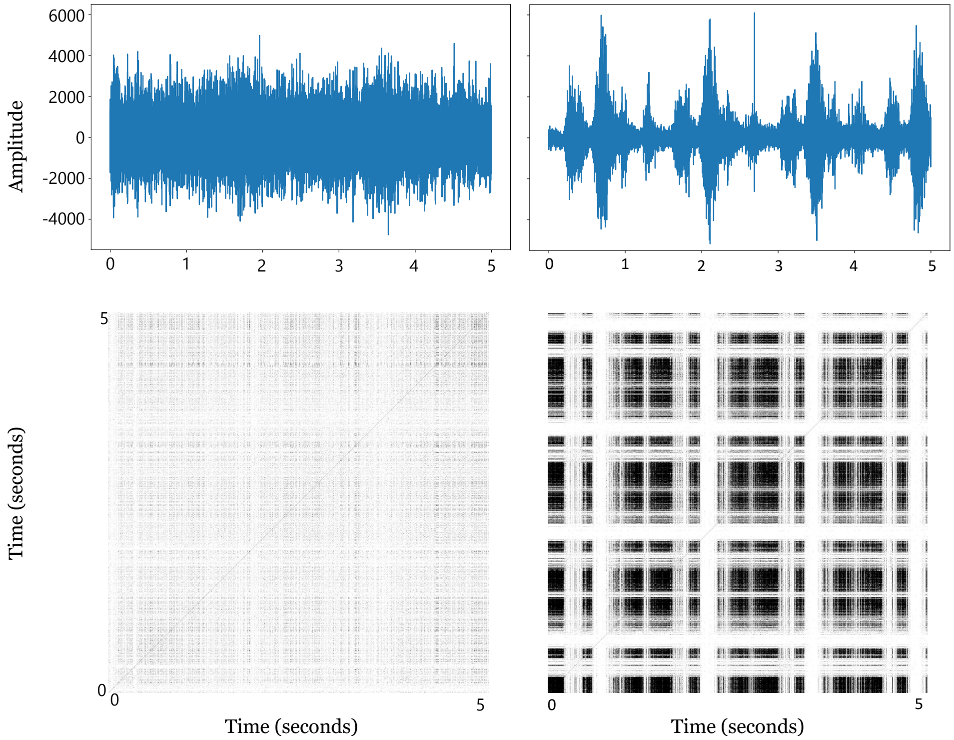
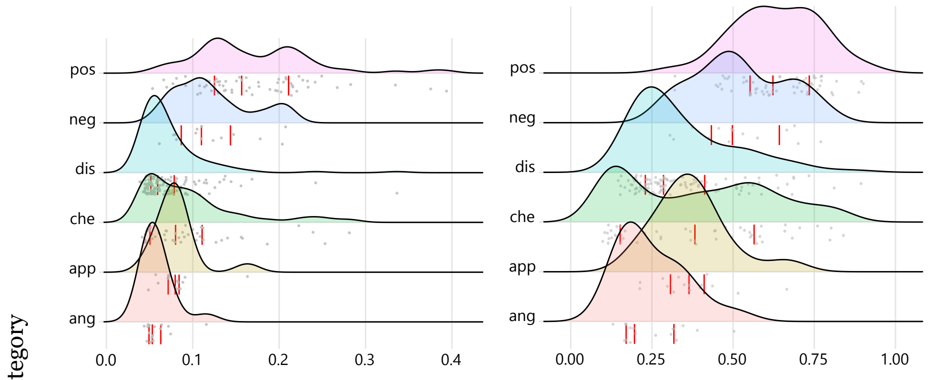


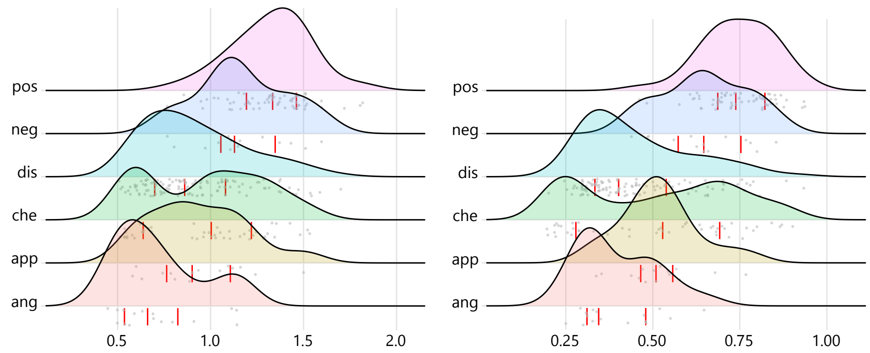
Figure 5.3: Representative timeseries and recurrence plots from five-second samples of two categories of crowd sound: Distraction Noise (left) and Positive Chant(right). During this sample of chant, the audience is repeating “De-fense, De-fense”.

Results of the linear regression with sum-to-zero contrasts are shown in Table 5.4. These results indicate that RQA metrics of most individual crowd sound categories can be differentiated from grand average RQA metrics across all crowd sound categories. At an alpha of 0.05, for the RQA metric of Laminarity, only Applause cannot be significantly differentiated from the grand average.Because positive chant was categorized as a reference level for the sum-to-zero contrast coding scheme, it’s comparison to grand average RQA metrics cannot be gathered from this model. Marginal means and pairwise comparisons with associated confidence intervals were extracted from the linear regression, and computed for each crowd sound category using R package emmeans, version 1.5.4 and are plotted in Figure 5.5. Blue bars represent 95 % confidence intervals, while red arrows represent comparisons among the means. Where a red arrow overlaps an arrow from another category, this means that the difference between the overlapping categories is not significant. These re-



a) Recurrence Rate

b) Determinism



c) Entropy

d) Laminarity

Figure 5.4: Ridgeline plots show the smoothed distribution and individual data points from the original (non-bootstrapped) distributions, along with quantile lines for four RQA Measures across six Crowd Sound Categories: Angry Noise, Applause, Cheer, Distraction Noise, Negative Chant & Positive Chant.

sults indicate that among most RQA metrics, Distraction Noise, Cheer, and Applause cannot be reliably distinguished from each other. Meanwhile, Angry Noise shows distinctly low values of each RQA metric, while both Positive and Negative Chant categories show distinctly high values of each RQA metric and are statistically different from each other in all metrics except for Entropy.

Table 5.3: Summary Statistics for each RQA Measure across six Crowd Sound Categories: Mean, median, standard deviation as calculated from original, non-bootstrapped data.

Recurrence Rate			
	mean	median	sd
AngryNoise	0.06	0.05	0.02
Applause	0.08	0.08	0.03
Cheer	0.10	0.08	0.06
DistractionNoise	0.07	0.06	0.04
NegativeChant	0.12	0.11	0.05
PositiveChant	0.17	0.16	0.07
Determinism			
	mean	median	sd
AngryNoise	0.25	0.20	0.10
Applause	0.37	0.36	0.11
Cheer	0.39	0.38	0.23
DistractionNoise	0.34	0.29	0.16
NegativeChant	0.52	0.50	0.14
PositiveChant	0.64	0.62	0.13
Entropy			
	mean	median	sd
AngryNoise	0.72	0.66	0.24
Applause	0.94	0.90	0.26
Cheer	0.96	1.00	0.32
DistractionNoise	0.92	0.86	0.27
NegativeChant	1.17	1.13	0.25
PositiveChant	1.32	1.33	0.21
Laminarity			
	mean	median	sd
AngryNoise	0.40	0.34	0.11
Applause	0.52	0.51	0.11
Cheer	0.51	0.53	0.22
DistractionNoise	0.45	0.40	0.16
NegativeChant	0.65	0.65	0.13
PositiveChant	0.75	0.74	0.10

Table 5.4: Results from the linear regression with sum-to-zero contrasts reported for each RQA measure across six crowd sound categories.

Recurrence Rate

	Estimate	Std. Error	t value	Pr(>|t|)
GrandMean(Int)	0.10	0.00	58.78	0.00
AngryNoise	-0.04	0.00	-10.97	0.00
DistractionNoise	-0.02	0.00	-6.69	0.00
Cheer	-0.02	0.00	-4.68	0.00
Applause	-0.01	0.00	-4.03	0.00
NegativeChant	0.03	0.00	8.33	0.00

R^2/R^2 adjusted 0.539/0.536

Determinism

	Estimate	Std. Error	t value	Pr(>|t|)
GrandMean(Int)	0.41	0.01	73.38	0.00
Angry Noise	-0.17	0.01	-13.59	0.00
DistractoinNoise	-0.07	0.01	-5.37	0.00
Cheer	-0.08	0.01	-6.31	0.00
Applause	-0.03	0.01	-2.62	0.01
NegativeChant	0.13	0.01	10.35	0.00

R^2/R^2 adjusted 0.525/0.522

Entropy

	Estimate	Std. Error	t value	Pr(>|t|)
GrandMean(Int)	0.99	0.01	100.97	0.00
AngryNoise	-0.30	0.02	-13.71	0.00
DistractionNoise	-0.07	0.02	-3.14	0.00
Cheer	-0.12	0.02	-5.64	0.00
Applause	-0.05	0.02	-2.27	0.02
NegativeChant	0.23	0.02	10.31	0.00

R^2/R^2 adjusted 0.510/0.507

Laminarity

	Estimate	Std. Error	t value	Pr(>|t|)
GrandMean(Int)	0.54	0.01	100.53	0.00
AngryNoise	-0.15	0.01	-12.64	0.00
DistractionNoise	-0.08	0.01	-7.05	0.00
Cheer	-0.08	0.01	-6.72	0.00
Applause	-0.01	0.01	-1.24	0.22
NegativeChant	0.13	0.01	10.56	0.00

R^2/R^2 adjusted 0.525/0.522

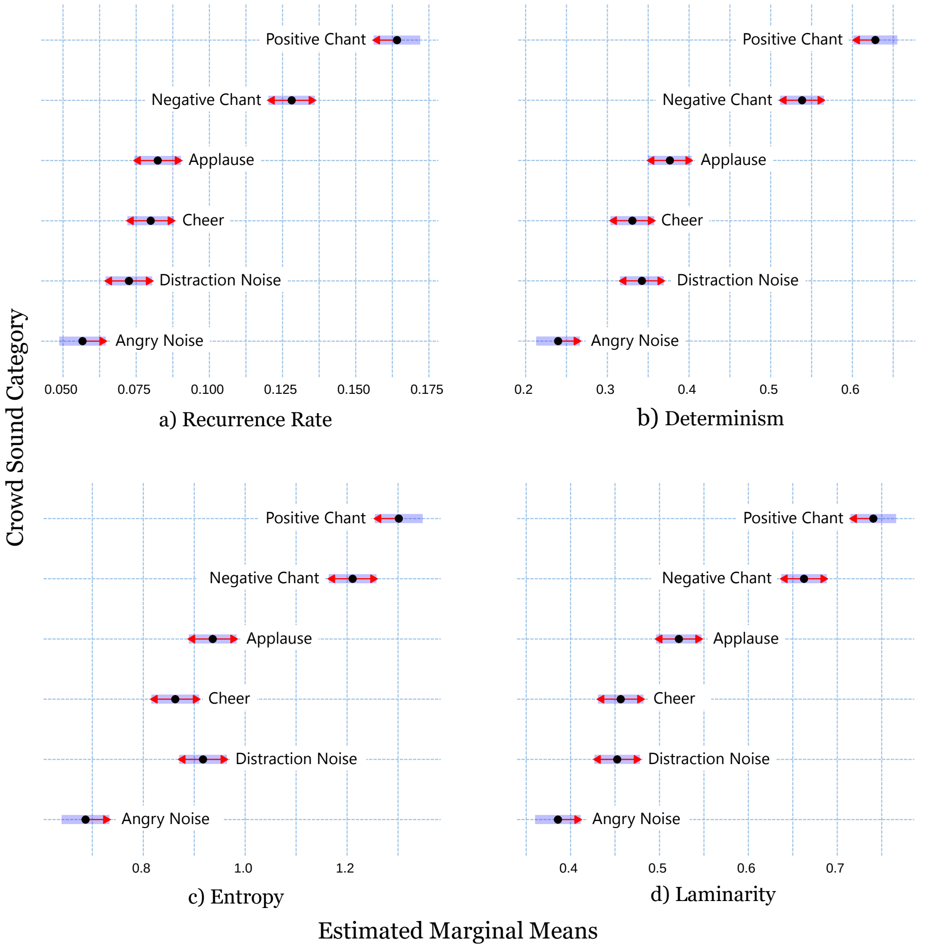


Figure 5.5: Pairwise Comparisons of Estimated Marginal Means across crowd sound categories for four RQA Metrics: A) Recurrence Rate, B) Determinism, C) Entropy, and D) Laminarity. Blue bars represent 95 % confidence intervals. Red arrows represent comparisons among the means. Where a red arrow overlaps an arrow from another category, the difference between the overlapping categories is not significant.

Table 5.5: Estimated Marginal Means.

Recurrence Rate				
	emmean	SE	lower.CL	upper.CL
AngryNoise	0.06	0.00	0.05	0.06
DistractionNoise	0.07	0.00	0.06	0.08
Cheer	0.08	0.00	0.07	0.09
Applause	0.08	0.00	0.07	0.09
NegativeChant	0.13	0.00	0.12	0.14
PositiveChant	0.16	0.00	0.16	0.17
Confidence level used: 0.95				

Determinism				
	emmean	SE	lower.CL	upper.CL
AngryNoise	0.24	0.01	0.21	0.27
DistractionNoise	0.34	0.01	0.32	0.37
Cheer	0.33	0.01	0.30	0.36
Applause	0.38	0.01	0.35	0.40
NegativeChant	0.54	0.01	0.51	0.57
PositiveChant	0.63	0.01	0.60	0.66
Confidence level used: 0.95				

Entropy				
	emmean	SE	lower.CL	upper.CL
AngryNoise	0.69	0.02	0.64	0.73
DistractionNoise	0.92	0.02	0.87	0.96
Cheer	0.86	0.02	0.82	0.91
Applause	0.94	0.02	0.89	0.98
NegativeChant	1.21	0.02	1.16	1.26
PositiveChant	1.30	0.02	1.26	1.35
Confidence level used: 0.95				

Laminarity				
	emmean	SE	lower.CL	upper.CL
AngryNoise	0.39	0.01	0.36	0.41
DistractionNoise	0.45	0.01	0.43	0.48
Cheer	0.46	0.01	0.43	0.48
Applause	0.52	0.01	0.50	0.55
NegativeChant	0.66	0.01	0.64	0.69
PositiveChant	0.74	0.01	0.72	0.77
Confidence level used: 0.95				

Part 2: SVM Classification with RQA features

The objective in Part 2 was to explore the usefulness of the full suite of 19 RQA metrics as features for the classifcation of different crowd sounds. To perform classification we utilized a support vector machine (SVM) classifier with an RBF kernel. The SVM classifier partitions the n-dimensional feature space (in our case $n = 19$) using hyper planes to best distinguish the data based on class. Samples extracted from the crowd sound data set for classification in Part 2 are different than the samples analyzed in Part 1. In order to create a larger data set from which to learn we took windowed samples from the entire length of each crowd sound event. Five-second samples were extracted sliding by one second windows at a time. This means a 7 second crowd event can generate 3 samples instead of a singular sample. In order to prevent over fitting, we used disjoint events for each class in the training, validation, and test sets. This means that even though we are windowing samples, there will be no overlap between distinct crowd events in the training, validation, and test sets. We then performed RQA analysis on each sample to generate the RQA features as described above. Whereas statistical analysis above focused on four RQA metrics commonly studied in behavioral experiments of human interaction, training and classification was computed on all 19 RQA metrics reported by PyRQA (Rawald et al., 2017):

- Minimum diagonal line length
- Maximum diagonal line length
- Minimum white vertical line length
- Recurrence rate
- Determinism
- Average diagonal line length
- Longest diagonal line length
- Divergence
- Entropy diagonal lines
- Laminarity
- Trapping time
- Longest vertical line length
- Entropy vertical lines
- Average white vertical line length
- Longest white vertical line length
- Longest white vertical line length divergence
- Entropy white vertical lines
- Ratio determinism/recurrence rate
- Ratio laminarity/determinism

In this analysis, we look at two different classification problems. In the first problem we look at 19 RQA metrics over six crowd sound classes defined in the analysis: angry noise, distraction noise, positive chant, negative chant, cheer, and applause. We also looked at a second classification problem after combining crowd sound classes that showed similar performance in the SVM classifier (trained on all 19 RQA metrics), and that showed overlap in the original means comparisons of the subset of four theoretically motivated RQA metrics described in Section 2.Thus we combined positive/negative chant into a singular chant class, cheer/applause into a singular cheer class, and kept distraction noise in a class by itself. We excluded angry noise from the conglomerated classes due to comparatively low prevalence in the data, and minimal confusion with crowd sound categories in the six-class SVM classifier results.

Table 5.6 lists the distribution of samples in the training set in each of the six original classes. Training samples were bootstrapped (over sampling the minority classes) so each class has 281 samples. Testing samples were not bootstrapped. Table 5.7 lists the distribution of samples in the training set by conglomerated class. As above, training samples were bootstrapped (over sampling the minority classes) so each class has 281 samples. Testing samples were not bootstrapped.

Table 5.6: Overlapping 5s samples used for RQA and SVM Classification

Crowd Sound Category	**# of samples**
Angry Noise	37
Applause	19 (smallest class)
Cheer	97
Distraction Noise	281 (largest class)
Negative Chant	26
Positive Chant	142

Table 5.7: Overlapping 5s samples used for SVM Classifcation after combining crowd sound classes

Conglomerated Category	**# of samples**
Applause/Cheer	116
Distraction Noise	281
Chant	168

Classification Results

Results of the SVM classifier performance on the test set data are displayed in the confusion matrix in Figure 4. Here we have the true label on the y axis and the predicted label on the x axis. The values have been normalized across the true label, since each class contains a different number of samples because we did not bootstrap the test set. When analyzing the results of the SVM trained on size classes we see strong distinction (above 0.8) for angry noise, distraction noise, positive chant, and cheer. We see confusions for applause with miss-classifications split between cheer and angry noise. We also see significant confusion for negative chant with a sizeable portion being classified as positive chant. Along with the statistical analysis performed earlier this led us to join classes into three subgroups: chant, cheer/applause, and distraction noise. When we train an SVM classifier on these joined classes we see performance above 0.8 for all classes. This demonstrates that features generated by RQA are distinct for the different classes and that we can use them to discern between different crowd events in classification problems.

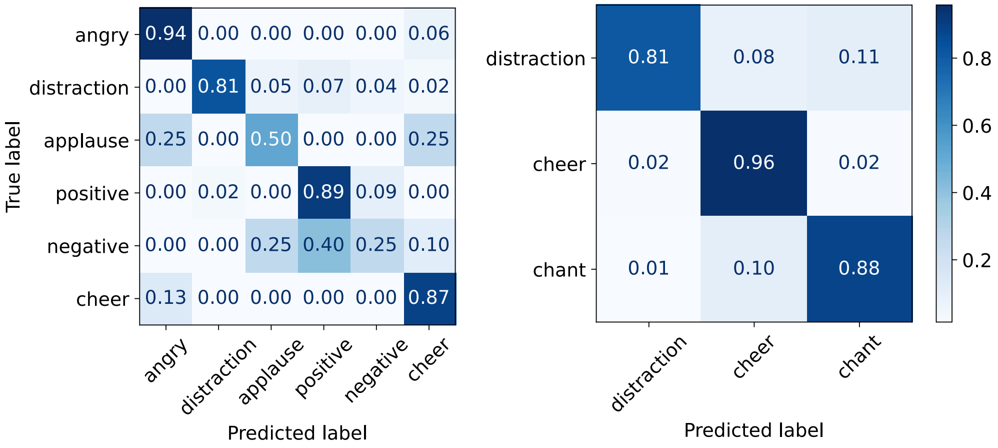


Figure 5.6: SVM test set results for both 6 classes (left) and 3 classes (right).

5.6 Discussion

We have empirically demonstrated differences in acoustic coordination patterns of differing crowd responses recorded from noisy, naturalistic acoustic social behavior of fans at a basketball game. The results here and from Proksch et al., 2022 (Proksch et al., 2022) indicate that RQA (and phase space reconstruction) can be meaningfully applied to global acoustic recordings when individual recordings are not available. The collegiate basketball crowd sound data set analyzed here is exemplar of naturalistic crowd interactions that are not dictated by a set of pre-determined instructions, like in a musical score. While crowd behavior is influenced by the events of the basketball game over time, the specific acoustic output of the crowd is neither rehearsed nor explicitly dictated via shared access to a ‘behavioral score’. Instead, individuals of the crowd ‘softly assemble’ into certain functional patterns acoustic of behavior, which emerge from local interactions among fans and influenced by external events of the environment. Further, we have demonstrated that a combination of RQA and SVM classifier can effectively differentiate between at least a subset of acoustic crowd responses. Future research involving classification of acoustic crowd behavior can test whether classification accuracy of SVMs, Naive Bayes Classifiers, or even Convolutional Neural Networks – which are typically trained on more standard measures of acoustic analysis (such as spectral features and mel cepstral coefficients) – may be improved by incorporating numerical RQA metrics or even Recurrence Plot images into the classification.

This work adds to the growing body of research on joint action and coordination among groups. Baranowski-Pinto et al 2022 found heart-rate interdependence between fans attending live (vs televised) basketball games, but did not measure synchrony in behavioral (i.e. movement or acoustic) dynamics. Swarbrick et al 2019 demonstrated that movement vigor and engagement was enhanced by attending a live concert compared to a pre-recorded concert, but did not analyse coordination in these dynamics between audience members. And Gordon et al 2020 demonstrated an independence between behavioral and physiological synchrony – such that physiological and behavioral synchrony are not always coordinated. Incorporating analysis of coordination measured from global acoustical signals (as shown in our work) to group interaction studies such as these will help shed light on the role of acoustical behavior in joint action, and whether that role is meaningfully correlated with movement, physiological, and psychological dynamics.

To conclude, we sought to describe emergent coordination dynamics in the acoustical behavior of a crowd in a naturalistic setting. Specifically, we performed phase space reconstruction and RQA on acoustic data recorded from fans attending a basketball game. While there was overlap in some categories, we found reliable differences in recurrence measures after SVM classification for three conglomerated categories of crowd activity (e.g. chant, cheer, and distraction noise). In the future it would be beneficial to analyze how these recurrence measures extend to additional basketball games or to acoustical behavior of crowds at different sporting events. Analyzing these signals over the time course of a game may shed light into how joint acoustical behavior changes over time. Further, it would be insightful to relate these acoustical behavioral dynamics with coordination dynamics across modalities.

Acknowledgements

We thank Sean Warnick (Computer Science, Brigham Young University) for assistance in providing the basketball crowd sound dataset and for early discussions during this project.

This material is based upon work supported by the National Science Foundation under Grant No. DGE-1633722. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Contributions

S.P. devised the project, the main conceptual ideas, and led writing the manuscript. M.R. performed data cleaning and preparation. S.P. and M.R. performed the nonlinear analysis, S.P performed statistical analysis. M.R. performed machine learning analysis. K.G. and M.T. collected and provided the original dataset, and consulted on analysis methods and results. R.B. and C.K. verified the analytical methods and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Epilogue

General Discussion

In this dissertation, I have presented work on how the brain processes rhythm, as well as work on the rhythms that underlie the interpersonal social interaction of large groups of people.

In Chapter 1, I reviewed how the perception of musical rhythm relies on the interaction of motor, parietal, and auditory regions of the brain within a pathway called the dorsal auditory stream. This pathway enables predictions from the motor cortex to travel via the parietal cortex and inform auditory regions where in time the next musical beat will land, according to the Action Prediction for Auditory Simulation (ASAP) hypothesis (Patel & Iversen, 2014). I argued that cortical networks proposed by the ASAP hypothesis, in conjunction with subcortical networks proposed by the Gradual Audiomotor Evolution hypothesis (Merchant & Honing, 2014), provide a converging and more complete description of the role of the motor system in musical timing perception (Proksch et al., 2020). Additionally, these hypotheses can be integrated under the Active Inference framework of sensory processing, which posits that action and sensation are not two separable processes, but rather that the brain+body system actively solicits internal predictive models of incoming sensory stimuli in an ongoing process of prediction error minimization (Adams et al., 2013).

In Chapter 2, I applied non-invasive transcranial magnetic stimulation (TMS) to temporarily down-regulate neural activity within the dorsal auditory stream (the posterior parietal cortex) and a motor region outwith this pathway (the supplementary motor area) to probe the causal role of the motor system in the neural response to deviant rhythmic stimuli. This chapter presented preliminary evidence indicating that, while each of these cortical areas may play a role in beat perception, down-regulation of these areas may not have a substantial effect on beat and meter perception at the level of event related evoked neural responses recorded using EEG. However, due the small sample size of the preliminary dataset and substantial variability across participants' response to both the rhythmic stimuli and to brain stimulation, this question requires further research.

Bridging the divide between music cognition and interpersonal coordination, Chapter 3 introduces the concept of an interpersonal synergy. Measurement of interpersonal synergies can be indicative of shared social cognition: of joint participation in co-regulating multiple patterns of activity between two or more agents engaged in a social interaction (De Jaegher et al., 2010). Ideally, in the controlled setting of a lab, we can measure when an interpersonal synergy develops by correlating individual signals recorded from each individual within a group social interaction, e.g. electrodermal activity, movement dynamics, speech signals, or even neural activity. However, individual measurements such as these may not always be available or easy to obtain in very large, naturalistic social interactions.

In Chapters 4 and 5, I leveraged tools from dynamical systems theory to evaluate the coordination dynamics of large, multi-agent groups of people, where we did not have access to these individual level signals. The signal that we did have access to was the group-level acoustic signal generated by two different interacting groups: a musical ensemble, and fans at a collegiate basketball game. Specifically, in Chapter 4 I showed that Recurrence Quantification Analysis (RQA), when applied to only a single aggregate measurement – sound –, can reveal coordination patterns that differ during independent behavior, and after the formation of a single complex system during interdependent interaction of a musical ensemble (Proksch et al., 2022). The interactions of this musical ensemble provided a strong test case, as the musicians' behavior is both rehearsed and performed according to a musical score. The musical score provided a ground truth against which we were able to compare the results of these statistical techniques. As musicians co-created a shared acoustic social world, we found that they formed a larger complex system – an interpersonal synergy unified over all the musicians in the ensemble. In Chapter 5, I showed that these same techniques can be applied to a much 'noisier' interpersonal interaction, one that lacks the specific instructions of a musical score. Specifically, I showed that RQA can be used to differentiate categories of crowd sound behavior from fans attending a collegiate basketball game. This research is significant because it highlights that, even without a predetermined behavioral score, individuals in a crowd 'softly assemble' into functional patterns of interpersonal coordination that are detectable from their acoustical output.

General Conclusion

This dissertation has presented theoretical and experimental work spanning sensorimotor neuroscience, coordination dynamics, and complex systems. By beginning with the fruitful ground of music cognition and an enactive, embodied, multiscale approach to investigating human brain and behavior, I was able to ask – What networks in the brain are responsible for our perception of events in musical time? How do large groups of people coordinate their behavior over time?

Investigating these questions, this dissertation has presented work from two very different scales of measurement – from the neural processes involved in beat and rhythm perception, to the large-scale rhythmic patterns of large group interpersonal coordination. Each of these areas of research are related in their emphasis on the complex interaction of brain, body, and environment in perception and behavior. Musical phenomena, including pitch and rhythm, are built by organizing regularities in melodic and rhythmic events. The deeply embodied nature of our perception of musical rhythm relies in no small part on the predictive capacity of the motor system, and our active engagement with music allows us to move to the beat and even to perceive the beat when we are standing still. Music provides a phenomenal playground for investigating not only predictive processes in the brain, but also the coordination dynamics of interpersonal interaction. The regularities that enable us to perceive musical rhythm also enable us to coordinate action among a large group of individuals, extending our cognitive and behavioral processes among a collective interpersonal interaction. The balance between naturalistic and controlled interaction afforded by a musical score allows us to explore the use of new analysis paradigms to describe these social interactions. Finally, we can apply what we learn in these musical contexts to large group interactions 'in the wild', such as at sporting events when there is no strict coordinating behavioral score. Future research in music and social cognition should continue to take a multiscale approach, integrating findings across scales of measurement and of neural, physiological, and behavioral processes to evaluate both the neural and behavioral dynamics of our often musical and always rhythmic social life.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2015, November). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing*, *16*(4), 325–332. doi: 10.1007/s10339-015-0648-2

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643.

Albarracin, M., Constant, A., Friston, K. J., & Ramstead, M. J. D. (2021). A variational approach to scripts. *Frontiers in Psychology*, 3035.

al Haytham, I. (ca 1030; 1989). The optics of Ibn al-Haytham. Translated by A. I. Sabra.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization [version 2; peer review: 2 approved]. *Wellcome Open Res*, *4*(63). doi: 10.12688/wellcomeopenres.15191.1.

Alviar, C., Dale, R., Dewitt, A., & Kello, C. (2020, September). Multimodal Coordination of Sound and Movement in Music and Speech. *Discourse Processes*, *57*(8), 682–702. doi: 10.1080/0163853X.2020.1768500

Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L., & De Volder, A. G. (2017). Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. *European Journal of Neuroscience*, *45*(11), 1439–1450.

Ashe, J., & Bushara, K. (2014). The olivo-cerebellar system as a neural clock. *Neurobiology of Interval Timing*, 155–165.

Atherton, G., Sebanz, N., & Cross, L. (2019). Imagine all the synchrony: The effects of actual and imagined synchronous walking on attitudes towards marginalised groups. *PloS one*, *14*(5), e0216585.

Auksztulewicz, R., Myers, N. E., Schnupp, J. W., & Nobre, A. C. (2019). Rhythmic temporal expectation boosts neural activity by increasing neural gain. *Journal of Neuroscience*, *39*(49), 9806–9817.

Balasubramaniam, R., Haegens, S., Jazayeri, M., Merchant, H., Sternad, D., & Song, J.-H. (2021). Neural encoding and representation of time for sensorimotor control and learning. *Journal of Neuroscience*, *41*(5), 866–872.

Balasubramaniam, R., Riley, M. A., & Turvey, M. (2000, February). Specificity of postural sway to the demands of a precision task. *Gait & Posture*, *11*(1), 12–24. doi: 10.1016/S0966-6362(99)00051-X

Balasubramaniam, R., Wing, A., & Daffertshofer, A. (2004, September). Keeping with the beat: movement trajectories contribute to movement timing. *Experimental Brain Research*. doi: 10.1007/s00221-004-2066-z

Balconi, M., & Fronda, G. (2020). The dialogue between two or more brains: the "hyperscanning" for organization. *Frontiers in Psychology*, *11*, 598332.

Baranowski-Pinto, G., Profeta, V. L., Newson, M., Whitehouse, H., & Xygalatas, D. (2022). Being in a crowd bonds people via physiological synchrony. *Scientific reports*, *12*(1), 1–10.

Bastian, A. J. (2006). Learning to predict the future: the cerebellum adapts feedforward movement control. *Current opinion in neurobiology*, *16*(6), 645–649.

Bates, D. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. doi: 10.18637/jss.v067.i01

Blake, E. C., & Cross, I. (2015). The acoustic and auditory contexts of human behavior. *Current Anthropology*, *56*(1), 81–103.

Bouwer, F. L., Werner, C. M., Knetemann, M., & Honing, H. (2016, May). Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia*, *85*, 80–90. doi: 10.1016/j.neuropsychologia.2016.02.018

Brodal, H. P., Osnes, B., & Specht, K. (2017). Listening to rhythmic music reduces connectivity within the basal ganglia and the reward system. *Frontiers in neuroscience*, *11*, 153.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, *195*(6), 2417–2444.

Buhl, J., Sumpter, D., Couzin, D., Hale, J., Despland, E., Miller, E., & Simpson, S. (2006). From Disorder to Order in Marching Locusts. *Science*, *312*, 1402–1406. doi: 10.1126/science.1125142

Buonomano, D. V. (2014). Neural dynamics based timing in the subsecond to seconds range. *Neurobiology of interval timing*, 101–117.

Butler, B. A., Pedersen, K., Cook, M. R., Wadsworth, S. G., Todd, E., Stark, D., ... Warnick, S. (2018). Classifying crowd behavior at collegiate basketball games using acoustic data. In *Proceedings of meetings on acoustics* (Vol. 35).

Cameron, D. J., Bentley, J., & Grahn, J. A. (2015). Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. *Frontiers in Psychology*, *6*, 366.

Cannon, J. J., & Patel, A. D. (2021). How beat perception co-opts motor neurophysiology. *Trends in Cognitive Sciences*, *25*(2), 137–150.

Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F., & Viale, M. (2010). Scale-Free Correlations in Starling Flocks. *Proceedings of the National Academy of Sciences*, *107*(26). doi: 10.1073/pnas.1005766107

Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2019, December). Body sway reflects joint emotional expression in music ensemble performance. *Scientific Reports*, *9*(1), 205. doi: 10.1038/s41598-018-36358-4

Chang, A., Kragness, H. E., Wei, T., Bosnyak, D. J., Thiede, A., & Trainor, L. J. (2020). Body Sway Predicts Romantic Interest in Speed Dating. *Social Cognitive and Affective Neuroscience*, 185–192. doi: 10.1093/scan/nsaa093

Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017, May). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences*, *114*(21), E4134–E4141. doi: 10.1073/pnas.1617657114

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cerebral cortex*, *18*(12), 2844–2854.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008b). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of cognitive neuroscience*, *20*(2), 226–239.

Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence.* Oxford University Press.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181–204.

Clayton, M., Sager, R., & Will, U. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. In *European meetings in ethnomusicology.* (Vol. 11, pp. 1–82).

Connell, J. P., DiMercurio, A., & Corbetta, D. (2017). Dynamic systems theory. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (p. 1-8). New York: Springer International Publishing.

Coull, J., & Nobre, A. (2008, April). Dissociating explicit timing from temporal expectation with fMRI. *Current Opinion in Neurobiology*, *18*(2), 137–144. doi: 10.1016/j.conb.2008.07.011

Coull, J. T., Cotti, J., & Vidal, F. (2016). Differential roles for parietal and frontal cortices in fixed versus evolving temporal expectations: Dissociating prior from posterior temporal probabilities with fmri. *Neuroimage*, *141*, 40–51.

Coull, J. T., Vidal, F., & Burle, B. (2016). When to act, or not to act: that's the sma's question. *Current Opinion in Behavioral Sciences*, *8*, 14–21.

Crowe, D. A., Zarco, W., Bartolo, R., & Merchant, H. (2014). Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *Journal of Neuroscience*, *34*(36), 11972–11983.

Cummins, F. (2013). Joint speech: The missing link between speech and music? *PERCEPTA-Revista de Cognição Musical*, *1*(1), 17–17.

Cummins, F. (2020). The territory between speech and song: A joint speech perspective. *Music Perception*, *37*(4), 347–358.

Davis, T. J., Pinto, G. B., & Kiefer, A. W. (2017). The Stance Leads the Dance: The Emergence of Role in a Joint Supra-Postural Task. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.00718

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, *14*(10), 441–447.

de Lafuente, V., Jazayeri, M., & Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas. *Journal of Neuroscience*, *35*(10), 4306–4318.

Delorme, A., & Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9–21.

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012, February). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, *141*(1), 49–53. doi: 10.1037/a0023843

Demos, A. P., Chaffin, R., & Logan, T. (2017). Musicians body sway embodies musical structure and expression: A recurrence-based approach. *Musicae Scientiae*, *22*(2), 244–263. doi: 10.1177/1029864916685928

dos Santos, L., Barroso, J. J., Godoy, M. F. d., Macau, E. E., & Freitas, U. S. (2014). Recurrence quantification analysis as a tool for discrimination among different dynamics classes: The heart rate variability associated to different age groups. In *Translational recurrences* (pp. 125–136). Springer.

Drost, U. C., Rieger, M., Brass, M., Gunter, T. C., & Prinz, W. (2005). When hearing turns into playing: Movement induction by auditory stimuli in pianists. *The Quarterly Journal of Experimental Psychology Section A*, *58*(8), 1376–1389.

Drost, U. C., Rieger, M., & Prinz, W. (2007). Instrument specificity in experienced musicians. *Quarterly Journal of Experimental Psychology*, *60*(4), 527–533.

Durstewitz, D. (2003). Self-organizing neural integrator predicts interval times through climbing activity. *Journal of Neuroscience*, *23*(12), 5342–5353.

D'Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015, March). What can music tell us about social interaction? *Trends in Cognitive Sciences*, *19*(3), 111–114. doi: 10.1016/j.tics.2015.01.005

Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. (1987, November). Recurrence

Plots of Dynamical Systems. *Europhysics Letters (EPL)*, *4*(9), 973–977. (Publisher: IOP Publishing)

Engel, A. K., Friston, K. J., & Kragic, D. (2016). *The pragmatic turn: Toward action-oriented views in cognitive science* (Vol. 18). mit Press.

FitzGerald, T. H., Dolan, R. J., & Friston, K. (2015). Dopamine, reward learning, and active inference. *Frontiers in computational neuroscience*, 136.

Francke, M., Pogromsky, A., & Nijmeijer, H. (2020, February). Huygens' clocks: 'sympathy' and resonance. *International Journal of Control*, *93*(2), 274–281. (Publisher: Taylor & Francis) doi: 10.1080/00207179.2019.1590736

Friston, K. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual review of neuroscience*, *25*(1), 221–250.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815–836.

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

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PloS one*, *4*(7), e6421.

Friston, K. J., & Friston, D. A. (2013). A free energy formulation of music generation and perception: Helmholtz revisited. In *Sound-perception-performance* (pp. 43–69). Springer.

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network neuroscience*, *1*(4), 381–414.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., … Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS computational biology*, *8*(1), e1002327.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458.

Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Journal of Neuroscience*, *35*(45), 15187–15198.

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *Journal of Neuroscience*, *32*(5), 1791–1802.

Gallagher, S., Hutto, D., Slaby, J., & Cole, J. (2013). The brain as part of an enactive system. *Behavioral and Brain Sciences*, *36*(4), 421-22.

Gámez, J., Mendoza, G., Prado, L., Betancourt, A., & Merchant, H. (2019). The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLoS biology*, *17*(4), e3000054.

Giovannelli, F., Innocenti, I., Rossi, S., Borgheresi, A., Ragazzoni, A., Zaccara, G., … Cincotta, M. (2014). Role of the dorsal premotor cortex in rhythmic auditory–motor entrainment: a perturbational approach by rtms. *Cerebral Cortex*, *24*(4), 1009–1016.

Gordon, C. L., Cobb, P. R., & Balasubramaniam, R. (2018, November). Recruitment of the motor system during music listening: An ALE meta-analysis of fMRI data. *PLOS ONE*, *13*(11), e0207213. doi: 10.1371/journal.pone.0207213

Gordon, I., Gilboa, A., Cohen, S., & Kleinfeld, T. (2020). The relationship between physiological synchrony and motion energy synchrony during a joint group drumming task. *Physiology & Behavior*, *224*, 113074.

Goupil, L., Saint-Germier, P., Rouvier, G., Schwarz, D., & Canonne, C. (2020). Musical coordination in a large group without plans nor leaders. *Scientific Reports*, *10*(1), 1–14. doi: 10.1038/s41598-020-77263-z

Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of cognitive neuroscience*, *7*(4), 497–510.

Grahn, J. A., & Brett, M. (2007, May). Rhythm and Beat Perception in Motor Areas of the Brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. doi: 10.1162/jocn.2007.19.5.893

Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). fmri investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *Neuroimage*, *54*(2), 1231–1243.

Grahn, J. A., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage*, *47*(4), 1894–1903.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, *29*, 7540-7548.

Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral cortex*, *23*(4), 913–921.

Grube, M., Lee, K.-H., Griffiths, T. D., Barker, A. T., & Woodruff, P. W. (2010). Transcranial Magnetic Theta-Burst Stimulation of the Human Cerebellum Distinguishes Absolute, Duration-Based from Relative, Beat-Based Perception of Subsecond Time Intervals. *Frontiers in Psychology*, *1*. doi: 10.3389/fpsyg.2010.00171

Harrington, D. L., Zimbelman, J. L., Hinton, S. C., & Rao, S. M. (2010). Neural modulation of temporal encoding, maintenance, and decision processes. *Cerebral Cortex*, *20*(6), 1274–1285.

Hohwy, J. (2013). *The predictive mind.* OUP Oxford.

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*(1), 85–91.

Honing, H., Bouwer, F. L., Prado, L., & Merchant, H. (2018, July).

Rhesus Monkeys (Macaca mulatta) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis. *Frontiers in Neuroscience*, *12*, 475. doi: 10.3389/fnins.2018.00475

Hove, M. J., & Keller, P. E. (2010). Spatiotemporal relations and movement trajectories in visuomotor synchronization. *Music Perception*, *28*(1), 15–26.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social cognition*, *27*(6), 949–960.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005, January). Theta Burst Stimulation of the Human Motor Cortex. *Neuron*, *45*(2), 201–206. doi: 10.1016/j.neuron.2004.12.033

Huron, D. (2008). *Sweet Anticipation: Music and the Psychology of Expectation.* MIT Press.

Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content.* MIT press.

InagakuenWindOrchestra. (2009, February). *Welcome to the Imagination World.* (Inagakuen Wind Orchestra. Youtube, uploaded by kamoshitamoyashi. 17 Feb. 2011, `https://www.youtube.com/watch?v=-wJ9ZsgO3QI`)

Iversen, J. R., & Balasubramaniam, R. (2016, April). Synchronization and temporal processing. *Current Opinion in Behavioral Sciences*, *8*, 175–180. doi: 10.1016/j.cobeha.2016.02.027

Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition*, *134*, 232–244.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in cognitive sciences*, *12*(7), 273–280.

Jacoby, N., & McDermott, J. H. (2017, February). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, *27*(3), 359–370. (Publisher: Elsevier) doi: 10.1016/j.cub.2016.12.031

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54–75. doi: 10.1037/a0024208

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–621. doi: 10.1017/S0140525X08005293

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.

Kello, C. T., Bella, S. D., Médé, B., & Balasubramaniam, R. (2017, October). Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of The Royal Society Interface*, *14*(135), 20170231. doi: 10.1098/rsif.2017.0231

Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*, 57-78.

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., & Feldman, R. (2017, December). Brain-to-Brain Synchrony during Naturalistic Social Interactions. *Scientific Reports*, *7*(1), 17060. (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s41598-017-17339-5

Kirchhoff, M. D., & Kiverstein, J. (2020). Attuning to the world: The diachronic constitution of the extended conscious mind. *Frontiers in psychology*, 1966.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, *31*(5), 354–364.

Kiverstein, J., Miller, M., & Rietveld, E. (2019, July). The feeling of grip: novelty, error dynamics, and the predictive brain. *Synthese*, *196*(7), 2847–2869. doi: 10.1007/s11229-017-1583-9

Koelsch, S., Vuust, P., & Friston, K. (2019, January). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. doi: 10.1016/j.tics.2018.10.006

Kok, V. J., Lim, M. K., & Chan, C. S. (2016). Crowd behavior analysis: A review where physics meets biology. *Neurocomputing*, *177*, 342-362.

Koshimori, Y., Strafella, A. P., Valli, M., Sharma, V., Cho, S.-s., Houle, S., & Thaut, M. H. (2019). Motor synchronization to rhythmic auditory stimulation (ras) attenuates dopaminergic responses in ventral striatum in young healthy adults:[11c]-(+)-phno pet study. *Frontiers in Neuroscience*, *13*, 106.

Koshimori, Y., & Thaut, M. H. (2018). Future perspectives on neural mechanisms underlying rhythm and music based neurorehabilitation in parkinson's disease. *Ageing research reviews*, *47*, 133–139.

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in cognitive sciences*, *22*(10), 896–910.

Krueger, J. (2009). Enacting musical experience. *Journal of Consciousness Studies*, *16*(2-3), 98–123.

Krueger, J. (2013). Empathy, enaction, and shared musical experience: evidence from infant cognition. In T. Cochrane, B. Fantini, & K. R. Scherer (Eds.), *The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control.* Oxford University Press.

Ladinig, O., Honing, H., Haden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception*, *26*(4), 377–386.

Large, E. W., Herrera, J. A., & Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in systems neuroscience*, *9*, 159.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, *106*(1), 119.

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

Lopez-Calderon, J., & Luck, S. J. (2014). Erplab: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M., & Vuust, P. (2019). Weighting of neural prediction error by rhythmic complexity: A predictive coding account using mismatch negativity. *European Journal of Neuroscience*, *49*(12), 1597–1609.

Maes, P.-J. (2016, March). Sensorimotor Grounding of Musical Embodiment and the Role of Prediction: A Review. *Frontiers in Psychology*, *7*. doi: 10.3389/fpsyg.2016.00308

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

Mann, R. P., Faria, J., Sumpter, D. J., & Krause, J. (2013). The dynamics of audience applause. *Journal of The Royal Society Interface*, *10*(85), 20130466.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320–339.

Marwan, N. (n.d.). *Cross Recurrence Plot Toolbox for MATLAB, Ver. 5.22(R32.4).*

Marwan, N. (2011, April). How to avoid potential pitfalls in recurrence plot based data analysis. *International Journal of Bifurcation and Chaos*, *21*(04), 1003–1017. doi: 10.1142/S0218127411029008

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*, 237-329.

Marwan, N., & Webber, C. L. (2015). Mathematical and Computational Foundations of Recurrence Quantifications. In C. L. Webber & N. Marwan (Eds.), *Recurrence Quantification Analysis* (pp. 3–43). Cham: Springer International Publishing. (Series Title: Understanding Complex Systems) doi: 10.1007/978-3-319-07155-8_1

Marwan, N., Wessel, N., Meyerfeld, U., Schirdewan, A., & Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, *66*.

Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays*, *22*(1), 94–103.

Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive brain research*, *21*(2), 139–170.

Matthews, T. E., Witek, M. A., Lund, T., Vuust, P., & Penhune, V. B. (2020, July). The sensation of groove engages motor and reward networks. *NeuroImage*, *214*, 116768. doi: 10.1016/j.neuroimage.2020.116768

Matthews, T. E., Witek, M. A. G., Heggli, O. A., Penhune, V. B., & Vuust, P. (2019, January). The sensation of groove is affected by the interaction of rhythmic and harmonic complexity. *PLOS ONE*, *14*(1), e0204539. doi: 10.1371/journal.pone.0204539

Matyja, J. R., & Schiavio, A. (2013). Enactive music cognition: background and research themes. *Constructivist foundations*, *8*(3).

McEllin, L., Knoblich, G., & Sebanz, N. (2020, December). Synchronicities that shape the perception of joint action. *Scientific Reports*, *10*(1), 15554. doi: 10.1038/s41598-020-72729-6

McNeill, W. H. (2022). Keeping together in time. In *Keeping together in time.* Harvard University Press.

Merchant, H., Bartolo, R., Pérez, O., Méndez, J. C., Mendoza, G., Gámez, J., … Prado, L. (2014). Neurophysiology of timing in the hundreds of milliseconds: multiple layers of neuronal clocks in the medial premotor areas. *Neurobiology of interval timing*, 143–154.

Merchant, H., Harrington, D. L., Meck, W. H., et al. (2013). Neural basis of the perception and estimation of time. *Annu Rev Neurosci*, *36*(1), 313–336.

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? evidence for the gradual audiomotor evolution hypothesis. *Frontiers in neuroscience*, 274.

Meyer, L. (1956). *Emotion and Meaning in Music.* University of Chicago Press.

Michard, Q., & Bouchaud, J.-P. (2005). Theory of collective opinion shifts: from smooth trends to abrupt swings. *The European Physical Journal B-Condensed Matter and Complex Systems*, *47*(1), 151–159.

Misaki, M., Kerr, K. L., Ratliff, E. L., Cosgrove, K. T., Simmons, W. K., Morris, A. S., & Bodurka, J. (2021). Beyond synchrony: the capacity of fmri hyperscanning for the study of human social interaction. *Social Cognitive and Affective Neuroscience*, *16*(1-2), 84–92.

Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? a meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, *72*, 13–20.

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences*, *114*(42), E8913–E8921.

Nalepka, P., Kallen, R. W., Chemero, A., Saltzman, E., & Richardson, M. J. (2017, May). Herd Those Sheep: Emergent Multiagent Coordination and Behavioral-Mode Switching. *Psychological Science*, *28*(5), 630–650. doi: 10.1177/0956797617692107

Newen, A., De Bruin, L., & Gallagher, S. (2018). *The oxford handbook of 4e cognition*. Oxford University Press.

Noë, A., Noë, A., et al. (2004). *Action in perception*. MIT press.

Novembre, G., Knoblich, G., Dunne, L., & Keller, P. E. (2017). Interpersonal synchrony enhanced through 20 hz phase-coupled dual brain stimulation. *Social Cognitive and Affective Neuroscience*, *12*(4), 662–670.

Nozaradan, S., Schwartze, M., Obermeier, C., & Kotz, S. A. (2017). Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex*, *95*, 156–168.

Néda, Z., Ravasz, E., Bréchet, Y., Vicsek, T., & Barabási, A.-L. (2000). The sound of many hands clapping. *Nature*, *403*, 849–850.

O'Regan, J. K., & Noë, A. (2001, October). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–973. doi: 10.1017/S0140525X01000115

Pabst, A., & Balasubramaniam, R. (2018). Trajectory formation during sensorimotor synchronization and syncopation to auditory and visual metronomes. *Experimental Brain Research*, *236*(11), 2847–2856.

Pantaleone, J. (2002). Synchronization of metronomes. *American Journal of Physics*, *70*, 992–1000. doi: 10.1119/1.1501118

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological cybernetics*, *113*(5), 495–513.

Patel, A. D., & Iversen, J. R. (2014, May). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*. doi: 10.3389/fnsys.2014.00057

Paxton, A., & Dale, R. (2017, July). Interpersonal Movement Synchrony Responds to High- and Low-Level Conversational Constraints. *Frontiers in Psychology*, *8*, 1135. doi: 10.3389/fpsyg.2017.01135

Penhune, V. B., & Zatorre, R. J. (2019). Rhythm and time in the premotor cortex. *PLoS Biology*, *17*(6), e3000293.

Perez, O., Kass, R. E., & Merchant, H. (2013). Trial time warping to discriminate stimulus-related from movement-related neural activity. *Journal of Neuroscience Methods*, *212*(2), 203–210.

Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic prototypes across cultures: a comparative study of tapping synchronization. *Music Perception: An Interdisciplinary Journal*, *36*(1), 1–23.

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., ... Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical Transactions of the Royal Society B*, *376*(1835), 20200334.

Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music perception*, *2*(4), 411–440.

Proksch, S., Comstock, D. C., Médé, B., Pabst, A., & Balasubramaniam, R. (2020). Motor and predictive processes in auditory beat and rhythm perception. *Frontiers in Human Neuroscience*, 375.

Proksch, S., Reeves, M., Balasubramaniam, R., & Spivey, M. (2021, January). *Coordination Dynamics of Multi-Agent Human Interaction* (Tech. Rep.). PsyArXiv. (type: article) doi: 10.31234/osf.io/nz4g2

Proksch, S., Reeves, M., Spivey, M., & Balasubramaniam, R. (2022). Coordination dynamics of multi-agent interaction in a musical ensemble. *Scientific reports*, *12*(1), 1–14.

Rao, R. P. N., & Ballard, D. H. (1999, January). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi: 10.1038/4580

Rauschecker, J. P. (2018, January). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*, *98*, 262–268. doi: 10.1016/j.cortex.2017.10.020

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800–11806.

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., & Kotz, S. A. (2019, October). Rhythm in speech and animal vocalizations: a cross[U+2010]species perspective. *Annals of the New York Academy of Sciences*, *1453*(1), 79–98. doi: 10.1111/nyas.14166

Rawald, T., Sips, M., & Marwan, N. (2017). Pyrqa - conducting recurrence quantification analysis on very long time series efficiently. *Computers and Geosciences*, *104*, 101.108.

Reddy, V. R., Sinha, A., & Seshadri, G. (2013). Fusion of spectral and time domain features for crowd noise classification system. In *2013 13th international conference on intellient systems design and applications* (pp. 1–6).

Richardson, D. C., & Dale, R. (2005, November). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*(6), 1045–1060. (Publisher: John Wiley & Sons, Ltd) doi: 10.1207/s15516709cog0000$_2$9

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. (2007, December). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891. doi: 10.1016/j.humov.2007.07.002

Riley, M., Balasubramaniam, R., & Turvey, M. (1999, March). Recurrence quantification analysis of postural fluctuations. *Gait & Posture*, *9*(1), 65–78. doi: 10.1016/S0966-6362(98)00044-7

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal Synergies. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00038

Ross, J., Iversen, J., & Balasubramaniam, R. (2018). Dorsal premotor contributions to auditory rhythm perception: Causal transcranial magnetic stimulation studies of interval, tempo, and phase. *bioRxiv*, 368597.

Ross, J. M., & Balasubramaniam, R. (2014, August). Physical and neural entrainment to rhythm: human sensorimotor coordination across tasks and effector systems. *Frontiers in Human Neuroscience*, *8*. doi: 10.3389/fnhum.2014.00576

Ross, J. M., Iversen, J. R., & Balasubramaniam, R. (2016, November). Motor simulation theories of musical beat perception. *Neurocase*, *22*(6), 558–565. doi: 10.1080/13554794.2016.1242756

Ross, J. M., Iversen, J. R., & Balasubramaniam, R. (2018). The role of posterior parietal cortex in beat-based timing perception: a continuous theta burst stimulation study. *Journal of Cognitive Neuroscience*, *30*(5), 634–643.

Ross, J. M., Proksch, S., Iversen, J. R., & Balasubramaniam, R. (in review). Contributions of left and right parietal cortex for musical beat perception: replication and extension.

Ross, J. M., Warlaumont, A. S., Abney, D. H., Rigoli, L. M., & Balasubramaniam, R. (2016). Influence of musical groove on postural sway. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 308–319. doi: 10.1037/xhp0000198

Sarno, S., de Lafuente, V., Romo, R., & Parga, N. (2017). Dopamine reward prediction error signal codes the temporal evaluation of a perceptual decision report. *Proceedings of the National Academy of Sciences*, *114*(48), E10494–E10503.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038.

Schiavio, A., Maes, P.-J., & van der Schyff, D. (2021). The dynamics of musical participation. *Musicae Scientiae*, 1029864920988319. doi: 10.1177/1029864920988319

Schinkel, S., Dimigen, O., & Marwan, N. (2008, October). Selection of recurrence threshold for signal detection. *The European Physical Journal Special Topics*, *164*(1), 45–53.

Schirmer, A., Fairhurst, M., & Hoehl, S. (2021, January). Being 'in sync'—is interactional synchrony the key to understanding the social brain? *Social Cognitive and Affective Neuroscience*, *16*(1-2), 1–4. doi: 10.1093/scan/nsaa148

Schmidt, R., & O'Brien, B. (1997, September). Evaluating the Dynamics of Unintended Interpersonal Coordination. *Ecological Psychology*, *9*(3), 189–206. (Publisher: Routledge) doi: 10.1207/s15326969eco0903$_2$

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in cognitive sciences*, *11*(5), 211–218.

Schultz, W. (2007, June). Multiple Dopamine Functions at Different Time Courses. *Annual Review of Neuroscience*, *30*(1), 259–288. doi: 10.1146/annurev.neuro.28.061604.135722

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive neuroscience*, *5*(2), 97–118.

Shimizu, D. (2016, June). *Welcome to the Imagination World.* (`www.windrep.org/Welcome_to_the_Imagination_World`)

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003, April). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332. doi: 10.1037/0096-1523.29.2.326

Singh, P. P., & Rani, P. (2014). An approach to extract feature using mfcc. *IOSR Journal of Engineering*, *4*(8), 21–25.

Spivey, M. (2020). *Who You Are: the Science of Connectedness.* MIT Press.

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1811–1832. doi: 10.1037/a0014510

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., & Keller, P. E. (2013, July). Musical groove modulates motor cortex excitability: A TMS investigation. *Brain and Cognition*, *82*(2), 127–136. doi: 10.1016/j.bandc.2013.03.003

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic bulletin & review*, *10*(3), 630–637.

Swarbrick, D., Bosnyak, D., Livingstone, S. R., Bansal, J., Marsh-Rollo, S., Woolhouse, M. H., & Trainor, L. J. (2019). How live music moves us: head movement differences in audiences to live versus recorded music. *Frontiers in psychology*, *9*, 2682.

Swathi, H., Shivakumar, G., & Mohana, H. (2017). Crowd behavior analysis: A survey. In *2017 international conference on recent advances in electronics and communication technology (icraect)* (p. 169-178). doi: 10.1109/ICRAECT.2017.66

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L. S. Young (Eds.), *Dynamical systems and turbulence, warwick 1980. lecture notes in mathematics* (Vol. 898). Berlin, Heidelberg: Springer.

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017, June). Neural Entrainment to the Beat: The "Missing-Pulse" Phenomenon. *The Journal of Neuroscience*, *37*(26), 6331–6341. doi: 10.1523/JNEUROSCI.2500-16.2017

Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. *Annual review of neuroscience*, *24*, 631.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *Journal of Neuroscience*, *31*(10), 3805–3812.

Theiner, G., Allen, C., & Goldstone, R. L. (2010, December). Recognizing group cognition. *Special Issue on Extended Mind*, *11*(4), 378–395. doi: 10.1016/j.cogsys.2010.07.002

Thiel, M., Romano, M., Kurths, J., Meucci, R., Allaria, E., & Arecchi, F. (2002, October). Influence of observational noise on the recurrence quantification analysis. *Physica D: Nonlinear Phenomena*, *171*(3), 138–152. doi: 10.1016/S0167-2789(02)00586-9

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Tunçgenç, B., Travers, E., & Fairhurst, M. T. (2021). Leadership and tempo perturbation affect coordination in medium-sized groups. *Scientific reports*, *11*(1), 1–11. doi: 10.1038/s41598-021-81504-0

van der Weij, B., Pearce, M. T., & Honing, H. (2017). A probabilistic model of meter perception: Simulating enculturation. *Frontiers in psychology*, *8*, 824.

Vicary, S., Sperling, M., von Zimmermann, J., Richardson, D. C., & Orgs, G. (2017, July). Joint action aesthetics. *PLOS ONE*, *12*(7), e0180101. doi: 10.1371/journal.pone.0180101

Vlachos, I., & Kugiumtzis, D. (2010, Jul). Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E*, *82*, 016207.

von Zimmermann, J., Vicary, S., Sperling, M., Orgs, G., & Richardson, D. C. (2018, January). The Choreography of Group Affiliation. *Topics in Cognitive Science*, *10*(1), 80–94. doi: 10.1111/tops.12320

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music–brain responses to rhythmic incongruity. *cortex*, *45*(1), 80–92.

Vuust, P., & Witek, M. A. G. (2014, October). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.01111

Wallot, S., & Leonardi, G. (2018). Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r. *Frontiers in psychology*, 2232.

Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015, April). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, *06*. doi: 10.3389/fpsyg.2015.00313

Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson, M. J. (2018, January). Creating Time: Social Collaboration in Music Improvisation. *Topics in Cognitive Science*, *10*(1), 95–119. doi: 10.1111/tops.12306

Webber, C. L., & Marwan, N. (2015). *Recurrence Quantification Analysis: Theory and Best Practices*. Spinger.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed.). Springer.

Wiltermuth, S. S. (2012, January). Synchronous activity boosts compliance with requests to aggress. *Journal of Experimental Social Psychology*, *48*(1), 453–456. doi: 10.1016/j.jesp.2011.10.007

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

Winkler, I., Haden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009, February). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, *106*(7), 2468–2471. doi: 10.1073/pnas.0809035106

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.

Zhang, L., Liu, X., Zhang, X., & Bu, B. (2011). Audio Classification Algorithm Based on Nonlinear Characteristics Analysis. In *Proceedings of 2011 Asia-Pacific signal and information processing association annual summit and conference.*