Auditory-Sensorimotor Brain Function During Mental Imagery of Musical Pitch and Rhythm

Rebecca W. Gelding

B Sc (Advanced Mathematics; Honours 1st Class)

M Ed

Department of Cognitive Science,

ARC Centre of Excellence in Cognition and Its Disorders

Faculty of Human Sciences

Macquarie University, Sydney, Australia

This thesis is presented for the degree of Doctor of Philosophy (PhD)

March, 2019

Table of Contents

Abstract v
Statement of Originality vi
Related Presentations and Publications vii
Statement of Contributions x
Acknowledgements 1

Chapter 1: Introduction 3

Chapter 2: Mental Imagery of Music: Review of the Literature 7

2.1 Introduction 8
2.2 Definitions and Concepts 10
2.2.1 *Explicit Imagery* 13
2.2.2 *Implicit Imagery* 18
2.2.3 *Audiation* 20
2.2.4 *Inducing Musical Imagery* 21
2.3 Neural Mechanisms of Musical Imagery 23
2.3.1 *Brain Regions Implicated in Musical Imagery* 23
2.3.2 *General and Specific Networks* 25
2.3.3 *Auditory-Motor Interactions* 25
2.4 Neural Oscillations as an Index of Brain Function 27
2.4.1 *Beta-Band (13 – 30 Hz)* 28
2.4.2 *Alpha- / Mu- Band (8 – 12 Hz)* 32
2.4.3 *Neural Entrainment* 34
2.4.4 *Brain Stimulation Studies* 35
2.5 Individual Differences in Musical Imagery 37
2.6 Movement and Musical Imagery 39
2.7 Implications for the Real World 41
2.8 Considerations for Future Research 43
2.9 References 46

Chapter 3: The Pitch Imagery Arrow Task: Effects of Musical Training, Vividness, and Mental Control 64

3.1 Introduction 66
3.2 Materials and Methods 71
3.2.1 *Participants* 71
3.2.2 *Stimuli* 72
3.2.3 *Procedure* 77

3.3 Results 78
3.3.1 *Overall Accuracy and Reaction Times* 79
3.3.2 *Bucknell Auditory Imagery Scale* 80
3.3.3 *Strategy Use* 81
3.3.4 *Additional Imagery Performance Measures* 84
3.3.5 *Multiple Regression Analysis* 85
3.4 Discussion 90
3.5 Conclusions 96
3.6 References 99
Chapter 4: Applying Moden Psychometric Techniques to the Pitch Imagery Arrow Task 103
4.1 Introduction 105
4.2 Study 1: Exploratory Phase 107
4.2.1 *Materials and Methods* 107
4.2.2 *Results* 111
4.2.3 *Discussion* 113
4.3 Cognitive Process Model 114
4.4 Study 2: Calibration Phase 117
4.4.1 *Materials and Methods* 118
4.4.2 *Results* 120
4.4.3 *Discussion* 127
4.5 General Discussion 128
4.6 References 131
Chapter 5: Pitch Imagery Depends Upon Coordination of Auditory and Sensorimotor Brain Activity 138
5.1 Introduction 140
5.2 Materials and Methods 143
5.2.1 *Participants* 143
5.2.2 *Stimuli and Apparatus* 144
5.2.3 *Data Acquisition and Analysis* 147
5.3 Results 154
5.3.1 *Behavioural Results* 154
5.3.2 *MEG Results* 156
5.4 Discussion 165
5.4.1 *Modulation of β-Band During Pitch Imagery* 166

5.4.2 *β-band Communication* 168
5.4.3 *Directed Connectivity of Auditory and Sensorimotor Cortices* 169
5.5 Conclusions 171
5.6 References 173
Chapter 6: Sensorimotor Contributions to Mental Imagery of Musical Rhythm 179
6.1 Introduction 181
6.2 Study 1: Behavioural Study 185
6.2.1 *Materials and Methods* 185
6.2.2 *Results* 191
6.2.3 *Discussion* 196
6.3 Study 2: MEG Study 197
6.3.1 *Materials and Methods* 197
6.3.2 *Results* 202
6.3.3 *Discussion* 210
6.4 General Discussion 213
6.5 References 216
Chapter 7: General Discussion 221
7.1 Summary of Findings 223
7.2 Contributions to the literature 230
7.2.1 *Validated Test of Pitch Imagery* 230
7.2.2 *Auditory-Sensorimotor Interactions* 231
7.2.3 *Individual Differences* 233
7.2.4 *Movement and Musical Imagery* 234
7.3 Conclusions 235
7.4 References 237

Abstract

This thesis examined how the auditory and sensorimotor regions of the human brain function and interact to support mental imagery of musical pitch and rhythm. Three sets of experiments were designed to: (1) develop and validate novel musical imagery tasks that address some prominent drawbacks of previous paradigms, and are suitable for use in magnetoencephalography (MEG) neuroimaging studies; (2) measure auditory-sensorimotor brain function during mental imagery of musical pitch; (3) measure auditory-sensorimotor brain function during mental imagery of musical rhythm. Pitch imagery was studied using a Pitch Imagery Arrow Task (PIAT). Behavioural testing showed that the task reliably induced imagery in participants with a range of musical training and that performance was correlated with scores on an established psychometric test of imagery ability (BAIS). Item Response Theory analysis showed that the task requires the ability to both maintain and manipulate pitches and overcome perceptual biases. Rhythm imagery was studied using a Rhythm Imagery Task (RIT), and performance was found to improve after short-term tapping training. Individual differences in auditory imagery vividness and mental control were found to be more important for pitch than rhythm imagery. During both pitch and rhythm imagery, MEG beta-band activity was modulated in a similar manner in both auditory and sensorimotor source regions, indicating coordination of activity between these regions. Individual differences in imagery ability were related to greater sensorimotor to auditory directed connectivity. Finally, short-term motor training modulated the amount of right sensorimotor activity during rhythm imagery. These results support the interpretation that musical imagery is associated with coordinated activity in these regions; that the left cerebral hemisphere plays a dominant role in pitch imagery manipulation and that sensorimotor activity in the right cerebral hemisphere is particularly important for tracking rhythm. These contributions to knowledge have implications for music education and for therapeutic interventions for disorders of hearing, memory or motor function.

Statement of Originality

I, Rebecca Gelding, certify that the work in this thesis entitled "Auditory-Sensorimotor Brain Function During Mental Imagery of Musical Pitch and Rhythm" has not been previously submitted for a degree nor has it been submitted as part of the requirements for a degree to any university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my work, and the preparation of the thesis itself has been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

The research presented in this thesis was approved by the Macquarie University Human Research Ethics Committee, reference number: **5201300453.**

Signed:

Rebecca W Gelding (Student ID: 41424484)

October, 2018

Related Presentations and Publications

Research reported in Chapter 3 has been presented:

- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2013, November). *A novel behavioural study of pitch imagery and perception.* Poster session presented at the 4th Australasian Cognitive Neuroscience Conference (ACNC), Monash University, Melbourne.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2013, December). *Mental imagery of musical pitch.* Paper presented at the MEG day, Swinburne University of Technology, Melbourne.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2014, March). *A novel method for inducing mental imagery of musical pitch*. Paper presented at the Music, Cognition and Action Symposium, MARCS Institute, University of Western Sydney, Sydney.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2016, June). *Introduction to the pitch imagery arrow task (PIAT).* Invited colloquium at the Music Mind and Brain Group, Goldsmiths, University of London, UK.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2016, August). *The pitch imagery arrow task: Effects of musical training, vividness and mental control.* Paper presented at the ARC Centre of Excellence in Cognition and its Disorders Annual Workshop, The Fairmont Resort, Leura.

Research reported in Chapter 4 has been presented:

- **Gelding, R.W.,** Harrison, P.M.C., Johnson, B.W., Thompson, W.F, & Müllensiefen, D. (2018, July). *Developing a psychometrically advanced version of the Pitch Imagery Arrow Task*. Paper presented at the 15th International Conference on Music Perception and Cognition (ICMPC15), Graz, Montreal, Sydney.

Research reported in Chapter 5 and presented:

- **Gelding, R.W.,** Thompson, W. & Johnson, B.W., (2014, August). *Manipulating music in the mind: Investigation of pitch imagery*. Poster session presented at the

ARC Centre of Excellence in Cognition and its Disorders Annual Workshop, Macquarie University, Sydney.

- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2015, April). *Music in the mind: A MEG study of pitch imagery.* Poster session presented at the 42nd Annual Australasian Experimental Psychology Conference (EPC), University of Sydney.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2015, September). *A MEG study with a validated pitch imagery task*. Symposium conducted at the Workshop on Music, Mind and Emotion, The University of New South Wales, Sydney.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2015, December). *Mental imagery of musical pitch: A MEG study with a validated mental imagery task*. Paper presented at the 2nd Conference of the Australian Music & Psychology Society, Western Sydney University, Sydney.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2016, June). *Manipulating music in the mind: Auditory brain oscillations during the perception and imagery of musical pitch.* Invited colloquium at the Speech Science Forum, University College London, UK.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2016, June). *Manipulating music in the mind: Auditory brain oscillations during the perception and imagery of musical pitch.* Invited colloquium at the Institute of Biomagnetism and Bio-Signal Analysis, University of Munster, Munster, Germany.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2016, June). *Auditory brain oscillations of musical imagery*. Poster session presented at the 22nd Annual Meeting of the Organisation for Human Brain Mapping (OHBM 2016), Geneva, Switzerland.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2016, July). *Auditory brain oscillations during the perception and imagery of musical pitch.* Paper presented at the 14th International Conference of Music Perception and Cognition (ICMPC14), San Francisco, USA.

Research reported in Chapter 6 has been presented:

- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2017, December). *Auditory-motor oscillations during mental imagery of musical rhythms*. Paper presented at the CCD-KIT MEG Workshop, Macquarie University.
- **Gelding, R.W.,** Thompson, W.F., & Johnson, B.W. (2017, December). *Rhythm imagery and short-term motor training.* Paper presented at the 3rd Conference of the Australian Music & Psychology Society, The University of Queensland, Brisbane.
- **Gelding, R.W.,** Thompson, W., & Johnson, B.W. (2018, July). *Effect of short-term motor training on rhythm imagery accuracy.* Paper presented at the 15th International Conference on Music Perception and Cognition (ICMPC15), Graz, Montreal, Sydney.

Publications:

The research reported in Chapter 3 has been published in:

- **Gelding, R.W.,** Thompson, W.F., Johnson, B.W. (2015). The Pitch Imagery Arrow Task: Effects of musical training, vividness, and Mental Control. *PLoS ONE* 10(3): e0121809. doi:10.1371/journal.pone.0121809

A demonstration of the PIAT task validated in Chapter 4 is available online:
http://shiny.pmcharrison.com/piat-demo/

Additional publications related to work in this thesis:

- **Gelding R.W.** & Sun Y. (2018). Commentary: Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced β-band oscillations during perception. *Frontiers in Neuroscience.* 12:50. doi:10.3389/fnins.2018.00050

Statement of Contributions

All elements of design, testing, analysis and writing are the work of the author except for the following contributions from collaborators:

Chapter 4

- Analysis of exploratory and calibration phase was completed by Daniel Müllensiefen and Peter M.C. Harrison.
- Implementation of the 3000 stimuli item bank (created by the author) in ShinyR application for online testing was completed by Daniel Müllensiefen and Peter M.C. Harrison.
- Testing of the 243 participants for the calibration phase was completed by Daniel Müllensiefen and Peter M.C. Harrison.

Each chapter contains a specific statement of authorship outlining the role the authors played.

Acknowledgements

I would like to sincerely thank my supervisor, Blake Johnson, for his unwavering support, patience, encouragement, insight and feedback. Taking on a part-time candidate is a long-term commitment; thanks for standing by me all these years. Thanks to Bill Thompson for his advice and comments that always helped me see the bigger picture, and for having me be part of your Music, Sound and Performance lab.

Thanks to Daniel Müllensiefen and Peter Harrison for being wonderful international collaborators – always happy to answer questions and pointing me in the right direction. I've learnt so much from both of you, and look forward continuing our work together.

Thanks to the funding bodies that have contributed to scholarships, prizes or grants that have enabled this research to be both completed and communicated: the Australian government, Macquarie University, ARC Centre of Cognition and Its Disorders, PLoS ONE ECR Travel Award and SEMPRE.

My sincere thanks to the many researchers whose discussions, feedback and general academic advice has helped this thesis in immeasurable ways; especially Andrea Halpern, Sophie Scott, César Lima, Claudia Lappe, Rebecca Schaefer, Sylvie Nozaradan and Virginia Penhune. Thanks also to Kate Stevens for mentoring advice that has always stuck with me.

I'm grateful to everyone who has helped answer my many questions about MEG acquisition, data analysis, and coding including: Wei He, Paul Sowman, Joann Tang, Leidy Ja, Valerie Pu, Liz Stylianou, David Meng, Robert Seymour and Benjamin Morillon. Thanks to Lesley Hughes for her administration help, particularly in applying for the Postgraduate Research Fund.

Thanks to the colleagues and friends who have journeyed alongside me and encouraged me along the way, especially: Vana Webster, Olivia Brancatisano, Yanan Sun, Robina Day, Amee Baird and Gemma Perry. I am particularly grateful to Anna Fiveash for her friendship, positivity, sense of humour, advice and inspiration. And of course, a special shout out to the #musicscience twitter community, especially members of the Australian Music Psychology Society, for being a vibrant and engaging group of generous researchers.

Thanks to all my family and friends for all their support and belief in me, especially to Mum and Stan for helping with child minding over the years.

To my kids, Alana and Oliver, who have only ever known me as a mum who studies: thank you for cheering me on, for your hugs and your laughter. As proud as I am of this thesis, I will always be more proud of you. Finally, I owe my deepest gratitude to my husband, Graham: you have sacrificed the most for me to do this, and have lovingly supported me through all the ups and downs. You are the best.

Chapter 1

Introduction

"*In auditory imagery lies one of the most precious of the gifts of music – the ability to live in a world of mental tones*" (Seashore, 1919, p. 211)

Musical imagery is a ubiquitous human experience, one that is important to musicians and non-musicians alike. It has implications for a number of real-world applications and is attracting increasing attention in scientific communities studying the psychology and neuroscience of music. The present thesis investigates the role of sensory and motor brain mechanisms in musical imagery; specifically, it used magnetoencephalography (MEG) to measure and characterise the function and interactions of auditory and sensorimotor brain regions while participants performed tasks designed to induce mental imagery of musical pitch and rhythm.

Chapter 2 provides a review of the literature and discusses concepts and terminologies relevant to mental imagery of music. The review considers methodological issues associated with the study of an inherently private experience like musical imagery, and individual differences associated with factors including musical training and imagery vividness. The neural mechanisms implicated in musical imagery by neuroimaging studies are reviewed. The chapter includes a consideration of the role that movement plays in musical imagery, and the real-world implications for this line of research. Finally, the chapter concludes with several recommendations for future studies of musical imagery.

Chapter 3 describes behavioural results from a novel musical pitch imagery task, the Pitch Imagery Arrow Task (PIAT). The PIAT was designed to address the shortcomings of previous musical imagery tasks and to reliably induce pitch imagery in participants with a range of musical abilities. The results of this chapter show that competent performance on the PIAT requires an active musical imagery strategy, and that self-reported auditory imagery vividness and mental control were more important that musical training in completing the task.

Chapter 4 provides a more detailed assessment of the psychometric properties of the PIAT using Item Response Theory (IRT). An exploratory study examined factors that contribute to item difficulty. A cognitive process model was then described to determine how participants were completing the PIAT. Finally, a calibration study systematically varied the main effects found in the exploratory phase to confirm which factors predict item difficulty. The results of the calibration study were modelled using IRT, and it was found that ability on the PIAT is best described as the ability to both maintain and manipulate pitches, as well as overcome perceptual biases in responding.

Chapter 5 implemented the PIAT in a MEG study which probed auditory and sensorimotor brain function while participants performed pitch imagery. The results showed significantly greater beta-band modulation during imagery compared to the control tasks of auditory perception and mental arithmetic. Imagery-induced beta modulation showed no significant differences between brain sources, which may reflect a tightly coordinated mode of communication between sensory and motor cortices via beta oscillations. Directed connectivity analysis revealed that, theta neural oscillations originated in the left sensorimotor region and were directed towards bilateral auditory regions during imagery onset as well as during perceptual temporal prediction. These results add to the growing evidence that motor regions of the brain are involved in the top-down generation of musical imagery, and that imagery-like processes may be involved in perception.

Chapter 6 validated a novel Rhythm Imagery Task (RIT) and then implemented this task in a MEG study which probed auditory and sensorimotor brain function while participants performed rhythm imagery. The task assessed the influence of rhythm imagery ability before and after short-term tapping training. In the initial behavioural study it was found that rhythm imagery performance was significantly better after

tapping. The MEG study confirmed that beta modulation in the right sensorimotor region tracked the metrical structure of downbeats. Directed theta band connectivity analysis revealed that accurate imagery performance was associated with increased sensorimotor to auditory connectivity. These results support the conclusion that successful RIT performance requires precise timing supported by the brain's motor systems.

Chapter 7 summarises the results of the thesis and considers their contribution to the literature. As well as the development of a new validated test of pitch imagery ability, and a new test of rhythm imagery, this thesis provides the first evidence that: (1) auditory-sensorimotor regions are coordinating during silent imagery of pitch and rhythm; (2) individual differences in imagery ability are related to perceptual processing; (3) auditory imagery vividness and mental control were more important for pitch imagery than for rhythm imagery; (4) short-term motor training improves subsequent rhythm imagery performance and modulates the amount of right sensorimotor activity in both rhythm imagery and perception.

Chapter 2

Mental Imagery of Music: Review of the Literature

Rebecca W. **Gelding**[a,c], William F. **Thompson**[b,c], Blake W. **Johnson**[a,c]

[a] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[b] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

[c] ARC Centre of Excellence in Cognition and its Disorders

Author Contributions

R.W.G. conducted the literature review and drafted the paper. All authors discussed the content and edited the manuscript.

Abstract

Musical imagery is the ability to represent music in the mind in the absence of sensory input. In the following, we survey the current state of knowledge in the field and review concepts and terminology. We then compare the various experimental methodologies for inducing musical imagery and consider the neural mechanisms that have been implicated by neuroimaging and neuropsychological studies. The putative role of movement in musical imagery is reviewed as well as the real-world implications for this research. The review concludes with a summary of gaps in the existing literature and recommendations for future studies.

2.1 Introduction

Mental imagery refers to the representation in the mind of a sensory experience in the absence of sensory input. It has been a topic of philosophical inquiry since antiquity -- Aristotle believed that imagination was central to thought itself (MacKisack et al., 2016). Modern investigations date back as early as 1880, when Galton used questionnaires to investigate mental imagery (Galton, 1880). Visual imagery became a topic of intense interest in modern cognitive psychology in the 1970's with the publication of Shepard and Metzler's seminal studies of mental rotation. In these studies, participants were presented with pairs of three-dimensional shapes at different orientations, and were required to decide if the shapes were the same or different ("parity judgment" task) (Shepard & Metzler, 1971). The results showed a smooth linear relationship between the time required to judge "correct" pairs, and the size of the angular difference or orientation between the pairs, indicating that participants performed the task by mentally rotating visual images in the same manner by which physical objects would be brought into alignment by a rotation through space.

It was a finding that "sent shock waves through cognitive psychology" (Corballis, 1997, p. 102), a field dominated at that time by computational models of cognitive processes: If parity judgements were performed by computerlike digital computations there would be no reason to expect that larger angular disparities would take longer to compute than smaller ones. The smooth, linear reaction time profiles suggested instead an *analogue* process by which template-like mental images were rotated through a trajectory. These results spoke directly to the underlying nature of mental representations and their transformations and initiated a rich period of mental imagery experimentation and theoretical debate that continues to the present day (Kosslyn, 1980, 1994; Pearson & Kosslyn, 2015; Pylyshyn, 1973, 1981).

The simple and elegant methodology of Shepard-Metzler's parity judgement task was a key innovation for mental imagery research, for several reasons. One is that the task was highly effective in inducing a common cognitive strategy in participants, as they reported performing the task by mentally rotating the shapes. The task is also difficult or impossible to perform successfully with alternative strategies. Moreover, similar reaction time profiles were obtained across different participants and across different groups of participants. Whatever the nature of the underlying mental process, parity judgments seemed to successfully target one with unique properties (e.g. a smooth reaction time profile) in isolation from others. Behavioural measurements were a second key innovation, providing rigourous and objective confirmation that participants were actually performing the task as instructed.

Visual imagery has dominated the field of mental imagery research and it was not until the 1990's that researchers turned their attentions to imagery in the auditory modality (Reisberg, 1992). Auditory imagery is defined as "the introspective persistence of an auditory experience, including one constructed from components drawn from long-term

memory, in the absence of direct sensory instigation of that experience" (Intons-Peterson, 1992, p. 46). Musical imagery is a subset of auditory imagery in which the "auditory experience" is a musical one. It has been described simply as the silent replaying of music in one's own mind (Halpern, 2003). The ability to hear music internally has been argued to be fundamental to musical talent (Gordon, 1989b; Seashore, 1919). The neural mechanisms of musical imagery have been studied since the 1990's, and steady progress was made over the next two decades using lesion studies (Zatorre & Halpern, 1993) neuroimaging techniques such as positron emission technology (Halpern & Zatorre, 1999; Zatorre, Halpern, Perry, Meyer, & Evans, 1996), electroencephalography (Janata, 2001a; Janata & Paroo, 2006) and functional magnetic resonance imaging (Halpern, Zatorre, Bouffard, & Johnson, 2004; Kraemer, Macrae, Green, & Kelley, 2005; Lotze, Scheler, Tan, Braun, & Birbaumer, 2003).

As neuroimaging techniques have improved, there has been a renewed interest in the research into the nature and neural mechanisms of musical imagery (Hubbard, 2010, 2018). The objective of this literature review is to synthesise the current concepts of musical imagery into a coherent framework, and to summarise current knowledge of the brain mechanisms implicated in musical imagery, firstly through review of the regions, networks and sub-networks of interest, and then through the brain oscillatory mechanisms involved in musical imagery. The role of movement in musical imagery is also discussed, as well as the real-world implications of this research. The reader is also directed to several recent articles for comprehensive reviews of musical imagery research (Halpern, 2003; Hubbard, 2010, 2013, 2018; Schaefer, 2017; Zatorre & Halpern, 2005).

2.2 Definitions and Concepts

As a field, mental imagery stands at the intersection of two key branches of cognitive psychology – perception and memory (Neisser, 1972). The interdisciplinary

nature of the research has led to the use of varied and confusing terminology, a problem that complicates any attempt to synthesise the extant literature on musical imagery (Godøy & Jørgensen, 2001; Schaefer, 2017). In the following, we discuss potential points of confusion and suggest an organising conceptual framework.

The notion that "imagery plays an essential role in perception" (Kosslyn, 1994, p. 145), that is, that perception is imagery-like, has been argued from the visual imagery literature, however less emphasis has been given to this idea in musical imagery literature, which has instead focused on how imagery is perception-like (Hubbard, 2010, 2013). One under-cited book, overlooked in all of the previously mentioned review articles on musical imagery, arose from the 1999 International Conference on Systematic and Comparative Musicology (Godøy & Jørgensen, 2001). Here, Janata (2001b) defines a theoretical relationship between imagery and perception with musical imagery in two specific contexts: (1) as a mental act internally generated from long-term memory, such as imagining a familiar melody ("*Non-expectant imagery*"), or (2) in the process of forming mental images while listening to music ("*Expectant imagery*"). For non-expectant imagery, Janata further distinguishes between an abstract image – e.g. a sequence of pitches in a melody devoid of specific instrumentation – and an eidetic image, a stronger sensory experience in which timbre of specific instruments is also experienced (Janata, 2001b).

A similar classification using different terminology was suggested by Moore (2011). Applying the types of mental imagery outlined by Strawson's philosophy of "Imagination and Perception" (Strawson, 1974) to music, Moore (2011) delineated four different types of musical imagery: sensory, constructive, creative and propositional. *Sensory imagery* is a deliberate and effortful bringing to mind of music, and is the type of imagery most studied experimentally (Schaefer, 2017). *Constructive imagery* is an

automatic function that enables the organisation of incoming information in order for the listener to understand it. It is driven by implicit internal representations allowing interpretation of perception, and may be the reason why there is so much apparent overlap between perceptual and imagery processes (Schaefer, 2014). *Creative imagery* refers to the imagination or creation of novel music in the mind; and *propositional imagery* refers to false beliefs and deliberately imagining that something else is the case (For more discussion see: Moore, 2011; Schaefer, 2014, 2017). The first two types of imagery outlined here (sensory and constructive) are similar to the distinctions made by Janata (2001b) (non-expectant and expectant). Hence both authors proposed that musical imagery is required for perceptual processing.

The distinction between sensory and constructive imagery can be more broadly defined and classified as a difference between *explicit* and *implicit* imagery (Figure 2-1); with sensory imagery describing the explicit experience of having a mental image of sound in the mind, whilst constructive imagery describes a more implicit experience of making sense of sensory information during music perception. The non-expectant imagery of Janata (2001b) fits into the "Voluntary Musical Imagery" section of this framework, while expectant imagery would fit both within the "Spontaneous Imagery" section and the implicit imagery sections. Schaefer (2017) argues that constructive imagery is needed in perception, and in line with predictive processing theories of perception (A. Clark, 2013), may account for the overlap in processing between imagery and perception. In order to conceptualise and synthesise the various terminology in the literature, each of the nodes of this framework are discussed below.

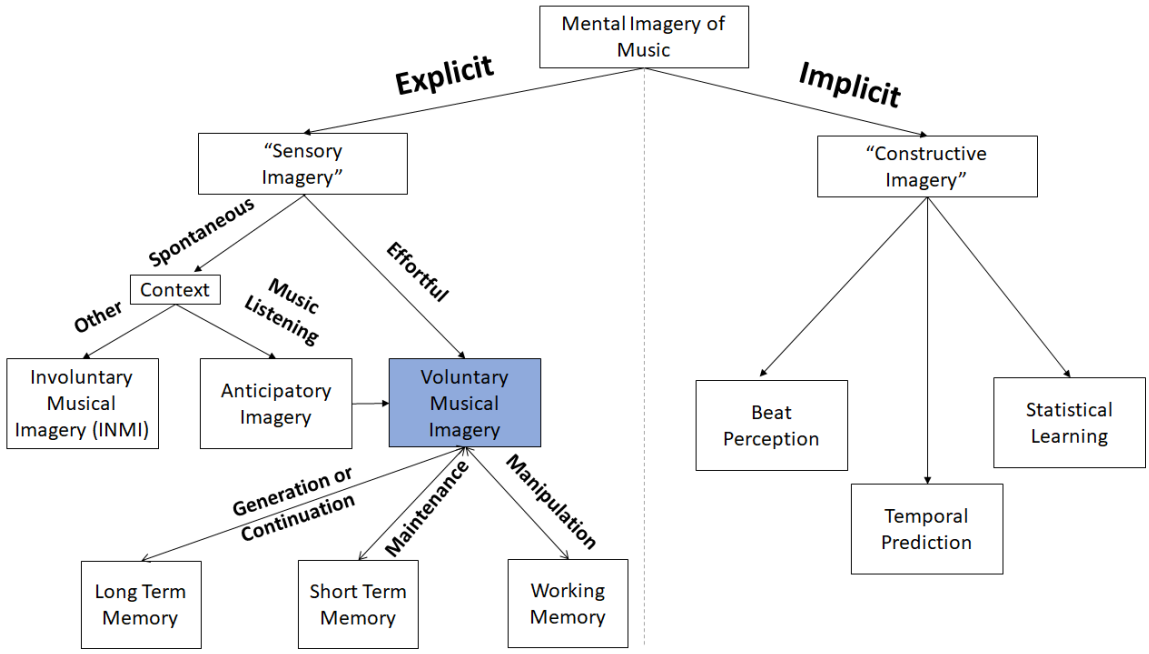

Figure 2-1: Terminology and concepts in mental imagery of music

2.2.1 Explicit Imagery

When the term "musical imagery" is used in the literature the most common distinction made is between voluntary and involuntary musical imagery (Weir, Williamson, & Mullensiefen, 2015). In line with Janata (2001b), the framework distinguishes Voluntary Musical Imagery as an effortful, explicit mental act. However, two subtypes of imagery that occur spontaneously, depending on the context, also occur explicitly. Anticipatory imagery is considered as a spontaneously initiated type of musical imagery that occurs in a music listening context, whereas involuntary musical imagery is a spontaneous cognition that occurs in contexts other than current music listening.

2.2.1.1 Spontaneous Imagery

Spontaneous imagery occurs in a variety of contexts. When the context is music listening, the term anticipatory imagery is often used as participants experience the musical image for the continuation of a musical stimuli. For example, when listening to a familiar song that has small segments replaced by white noise or silence, even under no

instruction to do so, participants report experiencing the continuation of the song through spontaneous imagery. This experimental paradigm has been used to compare neural correlates of imagery and perception using a variety of imaging modalities including functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and electroencephalography (EEG) (Gabriel et al., 2016; Kraemer et al., 2005; Müller et al., 2013). Recently, this type of paradigm has also been used to explore personal preference for a newly learnt song during imagery (Joucla et al., 2018).

Another form of spontaneous imagery occurs when listening to a well-known album, during the silence at the end of one song, where most people experience the beginning of the next song as a musical image in anticipation of hearing it (Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009). Similarly, anticipatory imagery appears to provide resolution to incomplete chord progressions (Otsuka, Tamaki, & Kuriki, 2008).

Importantly, the term anticipatory imagery has also been used to refer to the type of online musical imagery musicians use in anticipating actions and sound outcomes during musical performance (Keller, 2012; Keller & Appel, 2010). In the context of the proposed framework, this is more effortful rather than spontaneous imagery, and so this research falls within the "Voluntary Musical Imagery" category discussed below.

However not all spontaneous imagery occurs within a musical listening context. When melodies or fragments of songs spontaneously occur in the mind, that is not related to the current music listening context, they are labelled involuntary musical images (INMI), or "earworms" (Farrugia, Jakubowski, Cusack, & Stewart, 2015). The most common contexts in which INMI occur is after events such as recent music exposure or a memory trigger, or during certain internal states that are affective or require low attention (Williamson et al., 2012). While the tempo of songs spontaneously brought to mind, as well as voluntarily imagined under similar circumstances are similar

(Jakubowski, Bashir, Farrugia, & Stewart, 2018), people who report more involuntary musical imagery are not any better at voluntary musical imagery tasks (Weir et al., 2015). The term involuntary musical imagery can include non-pathological images, such as earworms, and auditory hallucinations associated with pathological conditions including head trauma, schizophrenia and deafness (Hubbard, 2018). Other spontaneous cognitions also occur in the context of music listening such as music-evoked visual imagery and mind-wandering (Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016; Taruffi, Pehrs, Skouras, & Koelsch, 2017).

2.2.1.2 Effortful Imagery

The topic of this thesis, Voluntary Musical Imagery (hereafter, "musical imagery"), is associated with conscious effort. Sound images, once generated by musical imagery, can be maintained and/or manipulated using long-term, short-term and working memory systems (Cowan, 2008). This is similar to the distinction that has been made in the visual imagery domain between image generation, maintenance and manipulation (Kosslyn, 1994).

Studies of musical imagery involving long-term memory representations have focused on either familiar melodies or on musical notes and scales. Some of the most frequently used musical imagery paradigms involved karaoke style tasks where lyrics appear on the screen and participants are required to mentally sing along in order to make judgements about pitches associated with specific words (Aleman, Nieuwenstein, Böcker, & de Haan, 2000; Halpern, 1992, 2003). Another approach is to play the start of a familiar instrumental melody, and require listeners to complete the melody using imagery, indicating when they reach the end with a button press (Halpern & Zatorre, 1999). A more rigorous design in recent studies of mental song completion, presents a single probe note after the completion phase and requires participants to judge whether the pitch or timing

of the probe is correct or incorrect (Herholz, Lappe, Knief, & Pantev, 2008, 2009; Weir et al., 2015). Some early imagery studies of musical notes, chords or scales simply instructed participants to imagine them, with no objective assessment of the occurrence or accuracy of the imagery (Meyer, Elmer, Baumann, & Jancke, 2007; Yoo, Lee, & Choi, 2001). More recently ascending scales were used, with the latter half of the scale needing to be imagined, and probe tones either mistuned or mistimed to varying degrees to measure the participant's imagery accuracy (Cebrian & Janata, 2010).

Maintenance (of musical imagery) paradigms draw heavily on short-term memory systems. The simplest maintenance paradigms expose participants to a sample tone sequence or melody and ask them to maintain these mentally while neurophysiological measures of brain activity (but not behavioural measures) are recorded (Schaefer, Desain, & Suppes, 2009; Schaefer, Vlek, & Desain, 2011b). Behavioural measures of imagery performance can be obtained with delayed match to sample (DMTS) designs. DMTS exposes participants to a sample of tones or short tone sequences, and requires them to hold these in memory and then decide whether or not a subsequent stimulus matches the sample (Deutsch, 1972; Grimault et al., 2014; Guimond et al., 2011; Lefebvre et al., 2013; Nolden et al., 2013; Schaefer et al., 2011b; van Dijk, Nieuwenhuis, & Jensen, 2010). A variation requires participants to mentally maintain a sample pattern and then to judge the pitch and timing of a probe stimulus (Kuchenbuch, Paraskevopoulos, Herholz, & Pantev, 2012). DMTS has also been used to measure imagery capacity, in terms of the amount of material that can be maintained for accurate performance for imagery of musical pitch and rhythm, and to explore the effects of brain stimulation on performance (Schaal, Banissy, & Lange, 2015; Schaal et al., 2014; Schaal, Pollok, & Banissy, 2017; Schaal, Williamson, et al., 2015).

Manipulation paradigms are designed to tax the working memory system. For example, participants are required to transpose simple melodic patterns (with no rhythmic variation) into different keys (Foster & Zatorre, 2010a, 2010b; Sutherland, Paus, & Zatorre, 2013), reverse them in time (Albouy, Weiss, Baillet, & Zatorre, 2017; Foster, Halpern, & Zatorre, 2013), or shift their serial order (Greenspon, Pfordresher, & Halpern, 2017). Mental reversal of both pitch and rhythm in melodies (e.g. Greensleeves), as has been used by Zatorre, Halpern, and Bouffard (2010), is a task that even experienced musicians find challenging. These studies have shown an important role for the intraparietal sulcus (IPS), part of the dorsal pathway, particularly in the right IPS for transposing melodies (Foster & Zatorre, 2010b), and the left IPS for reversing melodies in time (Albouy et al., 2017; Zatorre et al., 2010), though bilateral IPS activation is also seen for both types of manipulations (Foster et al., 2013; Foster & Zatorre, 2010a).

Finally, as mentioned previously, the majority of musical imagery research has focussed on how imagery is similar to perception by the exploration of different aspects of the music image (Hubbard, 2010, 2013). As well as imagery for pitch or melodic material discussed above, tempo imagery studies have revealed imagined tempos to be similar to perceived tempos (Halpern, 2003; Jakubowski, Farrugia, Halpern, Sankarpandi, & Stewart, 2015; Jakubowski, Farrugia, & Stewart, 2016). Loudness imagery has been found to be impaired with motor distractors, rather than alternative visual or auditory distractors; though musicians are able to overcome distractions better than non-musicians. These results suggest an important role in the motor system for imagery (Bailes, Bishop, Stevens, & Dean, 2012; Bishop, Bailes, & Dean, 2013, 2014). Imagery for timbre has also been explored (Halpern et al., 2004). Several studies have examined rhythm imagery by superimposing imagined accents – with varying metrical structure – onto a repeating rhythm (Fujioka, Fidali, & Ross, 2014; Iversen, Repp, & Patel, 2009; Nozaradan, Peretz,

Missal, & Mouraux, 2011; Schaefer, Vlek, & Desain, 2011a; Vlek, Schaefer, Gielen, Farquhar, & Desain, 2011). However, few studies have involved the imagination of rhythmic patterns in silence (Stupacher, Wood, & Witte, 2017).

2.2.2 Implicit Imagery

The three processes listed under constructive imagery in Figure 2-1 – beat perception, temporal prediction and statistical learning – are not exhaustive, but illustrative of the types of implicit imagery processes involved in music perception.

Beat perception has often been explored using syncopated rhythm stimuli which have no power at the beat frequency in the frequency spectrum of the sound envelope. Hence any enhanced brain responses at these frequencies can be argued to be endogenous or internally created (Nozaradan, Peretz, & Keller, 2016). This is evidence for a top-down and implicit construction of a beat on top of the heard stimulus, consistent with the notion that implicit imagery is required during beat perception. In a review of the beat perception literature Merchant, Grahn, Trainor, Rohrmeier, and Fitch (2015) showed the motor systems are involved in representations of musical structure of rhythmic patterns, and that these motor systems drive the auditory cortex with expectation for metrical stimuli, through beta and delta frequency bands.

However, when interpreting results of beat perception studies it is important to consider the behavioural task (if any). Chapin et al. (2010) investigated the role of attention in beat perception using complex syncopated patterns and simultaneous visually presented words, where participants had to rehearse either patterns or words for several seconds before responding either with rhythmic tapping (auditory condition) or verbal list of words (visual condition). The authors found that recruitment of the basal ganglia and SMA was enhanced during heard patterns in the auditory condition but not the visual

condition (Chapin et al., 2010). Interestingly, during silent rehearsal of auditory rhythms compared to words there was greater activity in bilateral basal ganglia, left medial prefrontal cortex, left post central gyrus and left primary auditory cortex. However, given the differences in the task outcomes between rhythmic tapping and verbally remembering words, it is difficult to determine if these differences in rehearsal period were due to the preparation of trial response. Using these same complex syncopated stimuli, Tal et al. (2017) had participants listen to the patterns for 32 seconds, before a probe presentation of a single pattern was presented, that was either on time or at a faster/slower tempo. Participants performed a tempo judgment on the probe pattern. This study found enhanced neural response, particularly in the right auditory cortex, at the frequency of the beat (Tal et al., 2017). Yet the required tempo judgement may have led to enhanced attention on tempo during the rhythmic pattern listening.

Temporal prediction is related to beat perception and has also been found to have motor origins (Morillon & Baillet, 2017; Morillon, Schroeder, & Wyart, 2014). Temporal prediction is explained by two complementary theories. Firstly, predictive coding posits that the brain is actively making predictions about incoming information and adjusting predictions based on inputs (A. Clark, 2013). Secondly, Dynamic Attending Theory is a framework that explains how perception of stimulus events is affected by their temporal context (Henry & Herrmann, 2014; Large & Jones, 1999). Temporal prediction occurs implicitly, and hence is an example of constructive imagery that is aiding perception.

Another constructive imagery phenomenon is statistical learning, proposed as a mechanism of language learning and argued to play a role in music perception and expectation (Loui, Wessel, & Kam, 2010; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010). Whilst musicians have been shown to be better statistical learners (Mandikal Vasuki, Sharma, Demuth, & Arciuli, 2016), a large online sample of the UK population

(N = 147,636) found active engagement in music related to musical sophistication, suggesting that deliberate music listening can train certain musical abilities in the absence of formal musical training (Mullensiefen, Ginras, Musil, & Stewart, 2014). Even through passive listening to novel tonal systems with artificial musical grammar, participants learn implicitly when expectations of the grammar are violated via statistical learning (Loui et al., 2010).

2.2.3 Audiation

The term "audiation" was first coined by Edwin Gordon and defined as "the hearing of music in one's mind when the sound is not physically present" (Gordon, 1985, p. 34). This definition then is synonymous with the common definition of "musical imagery" (Zatorre et al., 2010), however to Gordon, audiation was a broader concept entailing seven distinct types and five stages (Gordon, 1985); it was required to understand music that has just been heard, recall music, compose as well as perform. Gordon argues that "audiation is to music what thought is to language" (Gordon, 1999, p. 41), and that it is the central mental faculty that represents musical aptitude.

The stages of audiation outlined by Gordon feature a combination of conscious and unconscious mechanisms (Gordon, 1989b). He argues that while listening to music, listeners are unconsciously maintaining an "after-image" of what has just been heard (Stage 1) in order to integrate it into their conscious rhythmic and tonal structure for what has already been heard (Stage 2 and 3). He also argues that listeners are consciously predicting what they will hear next (Stage 5) (Gordon, 1985).

Gordon's theory of audiation is often overlooked in the current musical imagery literature. He argues that imagery involves explicit and implicit (or conscious and unconscious) processes. The notion that imagery-like processing is required in perception

are consistent with Gordon's model, albeit with less emphasis on whether listeners are conscious of the predictions they are making. Interestingly, Gordon's work resulted in the development of several tests (Primary / Intermediate / Advanced Measures of Musical Audiation) that music educators and researchers have used extensively to measure and predict future musical ability and aptitude (Mullensiefen et al., 2014; Puschmann, 2013; Schleuter, 1993). However, the audiation tests consist of same-different melodic discrimination tasks, which have been shown to involve a range of cognitive processes (P. M. C. Harrison, Musil, & Müllensiefen, 2016), and therefore are not specific enough to address individual differences.

2.2.4 Inducing Musical Imagery

As an inherently internal process, mental imagery can be a difficult area of research (Pearson, Naselaris, Holmes, & Kosslyn, 2015). One of the first hurdles to overcome in the study of any mental imagery, is adequate task design for inducing imagery in participants (Hubbard, 2013). In the study of musical imagery to date there are several shortcomings that arise in the literature which must be addressed in the design of new paradigms.

First, in some tasks there are no objective measures of performance; participants are simply asked to imagine a note or a song, and it is presumed that they have done that (Yoo et al., 2001). In other cases, participants are not even instructed to imagine the music, but are played familiar songs with gaps inserted, and afterward provide subjective confirmation that during the silence (or white noise) they experienced a continuation of a song (Kraemer et al., 2005; Müller et al., 2013). Subjective reporting is better than nothing but several authors argue it is important to have behavioural measures to test whether such imagery is actually occurring (Hubbard, 2010; Zatorre & Halpern, 2005).

Second, tasks are typically not adaptable to the abilities of individual participants. If tasks are too simple, then alternative cognitive strategies may be used rather than musical imagery. For example, if participants are required to imagine ascending musical scales and are given the first 5 notes of the scale, if they are only being testing on the final note of the scale then the tonality, or sense of key signature, may be used to complete the task rather than actively imagining the notes leading up to the tonic (Cebrian & Janata, 2010). On the other hand, some tasks such as the mental reversal of melodies are difficult even for accomplished musicians (Zatorre et al., 2010).

Third, many early paradigms involved a karaoke type task where words appeared on the screen to familiar songs and participants imagined parts of the songs. These types of tasks require participants to be familiar with specific songs and incorporate several aspects of the musical stimuli (pitch, contour, rhythm etc as well as language). To understand the contributions of pitch and rhythm separately, it is important to isolate these features.

Fourth, as Figure 2-1 suggests, musical imagery involves generation, maintenance and manipulation of a musical images, however most of the literature has focused on generation and maintenance. Of the manipulation paradigms that have been used, studies have investigated manipulation of melodies through transposition as well as mental reversal (Foster et al., 2013; Zatorre et al., 2010). However, given the difficulty of the reversal tasks, often only participants with substantial musical training are able to complete them (Zatorre et al., 2010). It remains unclear whether new paradigms requiring manipulation of single pitches rather than sequences, would be easier for non-musicians to complete.

Therefore new behavioural paradigms that wish to explore pitch and rhythm imagery separately should include objective measures of performance, be flexible in

design to be able to cater for a range of individual abilities and therefore be able to generalise the findings, and should confirm that a musical imagery strategy has been used to complete the task (Hubbard, 2018). Tasks that particularly investigate manipulation of musical imagery would be a valuable contribution to the literature.

2.3 Neural Mechanisms of Musical Imagery

Auditory imagery involves many, but not all, of the brain areas involved in auditory perception (Hubbard, 2010, 2013). Often discoveries in music perception have then led to investigations of whether these phenomena are also at work in imagery. Reflecting a change in focus of the neuroimaging field as a whole, music studies have shifted focus from individual brain regions to brain networks. For example, recent work has examined network interactions between auditory and sensorimotor regions to understand in more depth the nature of coordination between these regions in perception, particularly in beat perception and temporal prediction (Morillon & Baillet, 2017; Ross, Barat, & Fujioka, 2017; Tal et al., 2017).

2.3.1 Brain Regions Implicated in Musical Imagery

2.3.1.1 Auditory Cortex

Historically the study of mental imagery focused on whether the primary sensory areas were involved. In the visual domain, consensus has emerged that the primary visual cortex is activated during visual imagery (Kosslyn, 1973; Kosslyn & Thompson, 2003). In the auditory modality, the role of primary auditory cortex is less clear.

While some authors have reported activation of the primary auditory cortex during musical imagery (Kraemer et al., 2005) activation in secondary auditory cortex, particularly in the right hemisphere, is more reliably found (Halpern, 2003; Halpern & Zatorre, 1999; Herholz, Halpern, & Zatorre, 2012; Nolden et al., 2013; Otsuka et al.,

2008). In a recent study of a single musician undergoing epilepsy surgery, using electrocorticography (ECoG) to record directly from the brain, researchers found that specific neuronal assemblies in auditory regions were tuned to acoustic frequencies of the imagined music, whilst the subject played a well-known piece silently (Martin et al., 2017). A mismatch negativity response can also be elicited in musicians from an imagined trace of a familiar song when a wrong tone is reintroduced, suggesting that imagery and perception rely on a common neuronal network detecting expectancy violations within the auditory cortex (Herholz et al., 2008).

2.3.1.2 Motor Areas

Beyond the auditory cortex, passive music listening also involves motor regions of the brain (Chen, Penhune, & Zatorre, 2008; Zatorre, Chen, & Penhune, 2007). In studies of beat perception and temporal prediction, regions including supplementary motor area (SMA), premotor cortex, cerebellum and basal ganglia are recruited, even when participants are completely still (Chen et al., 2008; Grahn & Brett, 2007; Grahn & McAuley, 2009; Henry & Grahn, 2017).

These same motor regions are active during musical imagery. The activation of the SMA in early imagery tasks led researchers to conclude that musical imagery involves motor programs, such as imagined humming (Halpern & Zatorre, 1999). In a more recent review Lima, Krishnan, and Scott (2016) conclude that SMA and pre-SMA support the activation of sound-related motor representations in both perception and imagery. Hence imagery-like activation during perception may be due to common motor representation. In other types of auditory imagery, such as imagery of emotional words, content-specific representations could be successfully decoded in primary somatosensory and primary motor cortices (de Borst & de Gelder, 2016).

2.3.2 General and Specific Networks

Multimodal imagery studies and meta-analyses indicate that both modality-specific and cross-modal brain networks are involved in mental imagery (Daselaar, Porat, Huijbers, & Pennartz, 2010; McNorgan, 2012; Zvyagintsev et al., 2013). Domain-general imagery areas include bilateral dorsal parietal and left inferior frontal regions (McNorgan, 2012) and the default mode network (Daselaar et al., 2010; Zvyagintsev et al., 2013). In the auditory modality, imagery elicits activity in bilateral secondary auditory cortex, bilateral inferior frontal cortex, left supramarginal gyrus, left precentral gyrus and left SMA (McNorgan, 2012). Deactivation of a visual imagery specific network has been reported during auditory imagery, as well as vice versa (Zvyagintsev et al., 2013).

2.3.3 Auditory-Motor Interactions

A number of recent magnetoencephalography (MEG) studies have investigated the coordination of activities of auditory and motor brain networks in implicit imagery tasks including beat perception (Tal et al., 2017), temporal prediction (Morillon & Baillet, 2017) and passive listening (Ross et al., 2017).

Tal et al. (2017) had participants passively listen to complex syncopated rhythmic patterns, then make a tempo judgement on a probe presentation of the pattern, in order to determine the degree to which the beat was represented in auditory and motor brain regions. The authors found the right auditory cortex showed the biggest frequency response at the beat rate (2 Hz). The power at 2 Hz was also significantly negatively related to a subsequent behavioural measure of how long participants took to start tapping along to the complex patterns at the beat rate. Hence participants who had difficulty perceiving the beat (i.e. took longer to start tapping) did not show enhanced neural activity at the beat rate. The authors did not find significant beat related activity in motor areas,

as previously seen in the literature (Merchant et al., 2015) and argue while not contradicting the previous findings, it does leave open the question as to the exact nature of the dynamics of these regions during beat perception (Tal et al., 2017).

Bilateral auditory and sensorimotor regions of interest have also been studied in a temporal prediction task (Morillon & Baillet, 2017). In this study participants were required to separate two auditory streams at different temporal regularities, and judge whether the mean pitch height of a melody was higher or lower than the 4 isochronous tones used to introduce the tempo of the melody. The second auditory stream of distractor tones were played in antiphase to the tempo of the melody, such that the neural response to temporal prediction of the melody could be compared to the distractor tones. The authors used directed connectivity analysis and found significant right auditory to left sensorimotor directed connectivity during passive listening at the beat rate (3 Hz), and greater left sensorimotor to right auditory connectivity in the beta range (18 – 24 Hz) (Morillon & Baillet, 2017). The same patterns were shown in the left auditory to left sensorimotor connection, however the directed connectivity during passive listening in the right sensorimotor region of interest was not calculated. The passive listening condition was interspersed with a tapping condition in which participants tapped along to the frequency of the melody to help them attend to the pitches of the melody and ignore distractors. Participants completed two 1-hour sessions in the MEG on separate days. Behavioural performance during tapping was significantly better than passive listening (Morillon & Baillet, 2017). Yet the interspersed block design (20 trials of passive listening followed by 20 trials of tapping, repeated 5 times per session), may have influenced the auditory and sensorimotor connectivity, as the amount of sensorimotor engagement during passive listening has been found to be related to both short and long-term motor training (Krishnan et al., 2018; Ross et al., 2017). Hence it remains unclear whether

temporal prediction in passive listening would still show enhanced left sensorimotor to right auditory connectivity if the task had been completed with one session of passive listening and one session of tapping.

Finally, Ross et al. (2017) investigated auditory-sensorimotor interactions and the effect of short-term motor training within a passive listening paradigm. In this study, participants passively listened to an unfamiliar instrument (Tibetan singing bowl struck with a wooden mallet) for 24 min, then participants actively made the sounds themselves by striking the bowl for 30 min, followed by another passive listening period of 24 min. A control group had identical listening blocks, but during the sound-making period, initiated the sound of the bell through a button press on keypad rather than striking the bowl. Comparing the passive listening trials pre and post training, Ross et al. (2017) found greater event-related desynchronisation in the beta-band in right auditory and left sensorimotor sources, as well as increased functional connectivity in the theta band between the left sensorimotor source and bilateral auditory sources.

2.4 Neural Oscillations as an Index of Brain Function

When enough neurons in a small region of the brain fire together, their electrical current summates and can be detected on the scalp. Techniques such as electroencephalography (EEG) detect potential differences in the current between active electrodes and reference electrodes. Magnetoencephalography (MEG) is a complementary technique that measures the change in magnetic flux resulting from this electrical activity (Hari & Salmelin, 2012). The M/EEG time series signal can be thought of a summed series of sinusoidal waves with different frequencies. One analysis technique is to divide the time series into physiologically and behaviourally meaningful frequency bands, which have been shown to be relevant. The most common division of frequency bands is gamma (30 – 100 Hz), beta (15 – 30 Hz), alpha (8 – 12 Hz), and theta (4 – 8 Hz).

Once a frequency band of interest has been isolated the relative change in power at that frequency – either decreases or increases – are described as event related desynchronisation (ERD) or event related synchronisation (ERS) respectively (Pfurtscheller & Lopes da Silva, 1999).

The dynamics of the beta-band (15-30 Hz) in auditory-motor regions often resemble or parallel modulations in the alpha-band (~ 8-12 Hz) in that power decreases with active engagement (Spitzer & Haegens, 2017). For example, in an isochronous regular beat, when a tone is heard, there is a characteristic ERD of beta-band power typically 200 ms after the onset of the tone, which is thought to reflect exogenous processing (Fujioka, Trainor, Large, & Ross, 2012). The beta-band will then rebound back to its original state prior to the onset of the next expected tone, with the slope of the rebound adapting to the anticipated next beat (Fujioka et al., 2012). This beta rebound is thought to reflect endogenous components, related to top-down predictive sensory processing (Fujioka et al., 2012).

2.4.1 Beta-Band (13 – 30 Hz)

Historically the beta-band has been thought of as a motor rhythm (Cheyne, 2013) and has been implicated in motor imagery (Burianová et al., 2013). It was believed to reflect cortical idling, though more recent theories posit its role in maintaining the "status quo" in the sensorimotor regions (Engel & Fries, 2010). Others suggest that the beta-band reflects endogenous reactivation of cortical representations (Spitzer & Haegens, 2017).

Beta-band has also been implicated in perceptual prediction and timing in auditory regions (Fujioka et al., 2014; Fujioka et al., 2012), as well as in the auditory-motor interactions (Morillon & Baillet, 2017; Ross et al., 2017). It has been suggested that music cognition is dynamically embodied, with motor-action and perception representations

actively interacting (Maes, Leman, Palmer, & Wanderley, 2014) and that the beta-band is an "open-line" of communication between the auditory and sensorimotor regions (Tang, Crain, & Johnson, 2016), and a mechanism supporting auditory-motor connectivity (Henry & Grahn, 2017). If auditory-motor regions are involved in explicit imagery, then beta oscillations should be well-suited to serve as neural markers of these processes.

Explicit musical imagery studies using M/EEG have found modulations in beta-band activity. For example, Gunji, Ishii, Chau, Kakigi, and Pantev (2007) found that during imagined singing relative to a passive rest period, greater beta-band (15 – 30 Hz) ERD was seen in various motor regions including left SMA, right superior precentral gyrus, left inferior precentral gyrus, and left inferior postcentral gyrus. However no auditory regions showed significant changes in beta-band (Gunji et al., 2007). Yet this could be argued to be motor imagery activation rather than musical imagery, given the instruction for participants to explicitly imagine singing rather than imagine the music.

Several studies have investigated beta-band activity as a result of imagined accents providing different metrical structures to tones. When a downbeat is accented by physically increasing the volume of the tone a greater beta ERD is seen after the tone (Fujioka, Ross, & Trainor, 2015). Importantly, this same pattern is seen even when the accent is merely imagined rather than physically heard (Fujioka et al., 2015). Iversen et al. (2009) also showed increased beta power when participants imagined accents in non-isochronous rhythmic patterns.

Implicit imagery tasks investigating temporal and spectral prediction have also reported modulation of beta-band activity. Morillon and Baillet (2017) found that temporal prediction is encoded in interdependent delta (which was the stimulus rate) and beta-band oscillations in the left sensorimotor source and directed towards bilateral auditory sources. Using mismatch negativity paradigms with predictable or unpredictable

deviant tones Chang et al. (2016, 2018) have demonstrated that beta-band is also modulated by pitch predictability. Recently this group reported that greater beta-band desynchronisation is seen prior to a predictable deviant than an unpredictable one, and that on a trial-by-trial basis the greater the desynchronisation prior to the predictable deviant tone, the less the P3a response to the tone, which suggests a reduced attentional response to the tone. Hence they concluded that larger beta desynchronisation depth indicates predictability of upcoming deviant pitch (Chang, Bosnyak, & Trainor, 2018).

2.4.1.1 Theoretical Considerations for the Beta-Band

To date there is no unifying theory of cortical beta oscillations (Spitzer & Haegens, 2017) however several theories have been posited to account for beta-band activity in sensorimotor regions and imagery processes. For example, one of the main findings in the imagined accents of different metrical structure study of Fujioka et al. (2015) is that there was greater beta ERD in the auditory cortex for the downbeat relative to upbeats for both heard and imagined accents on the downbeat. Two potential theoretical explanations for this suggested by Teki and Kononowicz (2016) are predictive coding (Arnal & Giraud, 2012; A. Clark, 2013), in which the ERD is reflecting the anticipation of the next event to occur, and event tagging (Hanslmayr & Staudigl, 2014), in which the beta ERD is encoding the salient events that have previously occurred, such as a downbeat. If predictive coding were true, the ERD should be larger in the time window preceding the downbeat, however if event tagging were true the ERD should be larger in the epoch following the downbeat. Results of a recent mismatch negativity paradigm that maintained isochronous timing of presentation of tones, with either a regular or pseudorandom presentation of deviants, revealed greater ERD prior to an expected deviant, which the authors argue is evidence for predictive coding rather than event tagging (Chang et al., 2018). However, these theories are not mutually exclusive, and both

may be true. Hence the authors posit that lower beta-band (< 20 Hz) may be described by event tagging mechanisms whilst the whole of the beta-band (15 – 30 Hz) may be understood by predictive coding mechanisms (Chang et al., 2018). One other aspect that may distinguish the two theories is the role of attention in the task, as the MMN paradigm of Chang et al. (2018) had participants passively listen to tones while watching a silent movie, whereas active tasks requiring participants to maintain pitch or rhythm through imagery may show different mechanisms at work.

In addition the Action Simulation for Auditory Prediction theory (Patel & Iversen, 2014) posits that beat perception is a complex brain function involving temporally-precise communication between auditory and motor planning brain regions. The authors argue that motor planning regions simulate a periodic movement, which provides a neural signal, namely the beta-band, that helps the auditory system predict the timing of upcoming beats (Patel & Iversen, 2014). A testable prediction that arises from this theory is that the direction of beta-band activity in the auditory-motor subnetwork during beat perception is from motor to auditory regions. Their theory is consistent with inverse models of the dorsal auditory processing stream, that suggest perception involves input from motor to auditory cortex (Rauschecker, 2011; Rauschecker & Scott, 2009).

In summary, the theories described above have been specifically constructed from perceptual studies in which participants use implicit imagery to perceive a beat or predict an upcoming auditory stimulus. They suggest that the motor cortex is generating predictions of timing of events, communicating this information to the auditory cortex to aid in their perception. However the notion of directionality of beta-band activity between auditory-motor systems is not well understood (Arnal & Giraud, 2012). Recent investigations provide preliminary evidence that in the beta-band, the left sensorimotor region drives the right auditory region during temporal prediction (Morillon & Baillet,

2017). It is also unclear exactly how the modulation of the beta-band is tracking what has just occurred as well as what is expected to occur next, and when it is expected to occur.

2.4.2 Alpha- / Mu-Band (8 – 12 Hz)

There are several physiologically distinct brain rhythms in the alpha frequency band (8 – 12 Hz). The most prominent being the parieto-occipital alpha rhythm, and the rolandic mu rhythm localised to sensorimotor regions (Hari & Salmelin, 1997). There is also a temporal localised rhythm which has been described with various terminology including “tau rhythm” (Lehtelä, Salmelin, & Hari, 1997) or “auditory alpha rhythm” (Weisz, Hartmann, Müller, & Obleser, 2011). For the purposes of this review the term mu and alpha will be used interchangeably to refer to the 8 – 12 Hz frequency band activity in auditory and sensorimotor regions.

Studies using different paradigms have shown that the alpha / mu band also plays an important role in imagery of music. First, greater ERD in alpha-band (defined in this study as 8 – 15 Hz) in the motor region of left superior precentral gyrus (BA 4) is seen during imagined singing relative to a passive rest period (Gunji et al., 2007). Second, when listening to white noise inserted in both familiar and unfamiliar music, as participants report experiencing the illusion of hearing the missing familiar songs during the noise, alpha power reductions (ERD) were most prominently found in bilateral secondary auditory cortex, possibly also in the primary auditory areas (Müller et al., 2013). Though authors admitted they could not definitely state this from their data, they suggest it looked likely (Müller et al., 2013). Though it was not the focus of their study, Müller et al (2013) also note they found alpha power modulations in the inferior and medial frontal cortex and right precentral cortex.

Correlations between oscillatory power and behavioural performance have also been found. For example, both alpha and low beta (13.5 – 18 Hz) power were significantly negatively correlated with performance (accuracy/RT) in an EEG spatial / semantic auditory working memory task (Backer, Binns, & Alain, 2015). Better performance was associated with greater ERD in alpha and beta-bands. However, power was calculated from averaging the 15 electrodes for each frequency band that gave the largest desynchronisation during the encoding / maintenance phase. This may have led to biases in results.

In contrast, alpha-band activity has also been shown to increase during imagery compared to perception. When comparing the imagery and perception of short well known (overlearned) musical phrases, many individual differences were found in the topography of alpha-band activity, with an overall increase in alpha-band activity seen found in parieto-occipital regions (Schaefer et al., 2011b). Maintaining a tone in working memory in a delayed-matched-to-sample task, a left temporal increase in 5 – 12 Hz band activity is seen compared to a control condition with no active maintenance (van Dijk et al., 2010).

2.4.2.1 Alpha as Inhibition Hypothesis

While such increases in alpha may appear to contradict the earlier studies showing greater ERD, in fact the distributions of activity reveal that both are consistent with the alpha as inhibition hypothesis (Jensen & Mazaheri, 2010) – with decreases in alpha seen in auditory and motor regions, and increases in visual areas suggesting an active inhibition of visual processing during the internally directed activity of musical imagery.

2.4.3 Neural Entrainment

Neural entrainment occurs when neural oscillations synchronise with an external or exogenous stimulus (Thut, Schyns, & Gross, 2011), and typically occurs in the delta frequency range (0.5 – 4 Hz) in sensory cortices corresponding to a rhythmic beat rate of 250 ms - 2 seconds (Henry & Grahn, 2017). Nozaradan et al. (2011) has shown that imagining an accent of a downbeat in a binary or tertiary pattern on top of an isochronous pattern leads to changes in the entrainment of the brain to the beats, with peaks in the frequency spectrum ("steady state evoked potential": SSEP) occurring at downbeat frequencies, even though these downbeats are not physically accented. Further, using both unsyncopated and syncopated rhythm stimuli, the authors found that this reflects exogenous and endogenous processing of the beat, with syncopated rhythms requiring greater endogenous beat perception (Nozaradan et al., 2016). In light of the proposed framework (Figure 2-1), this supports the notion that beat perception during syncopated rhythms requires greater implicit imagery. The authors found that stronger endogenous neural entrainment at the beat frequency was associated with enhanced temporal prediction abilities and greater tapping accuracy in a subsequent tapping synchronisation task (Nozaradan et al., 2016). Using a similar frequency tagging approach to investigate the auditory-motor interactions, Tal et al. (2017) found that only the right auditory source showed significant increase in amplitude at the silent downbeat frequency in syncopated rhythms.

With a more realistic stimulus of high hat, snare drum and bass drum, Stupacher, Witte, Hove, and Wood (2016) had participants passively listen to repeating drum patterns, with one, two or three bars of silence inserted over the trial. They wanted to keep the drum break comparable to everyday music listening, and so participants were not explicitly instructed to try to imagine the drums during the silence. The results showed an

SSEP at the beat frequencies during listening but not silence. It was not reported if participants were asked if they imagined the drums during the silence, and so it remains unclear whether SSEPs would have been present had the participants been actively imagining the drums. In a follow up study comparing musicians and non-musicians, Stupacher et al. (2017) presented different 3 over 4 polyrhythms with drum kit and wood block playing separate rhythms. Silence was then inserted for at least one bar, and during the second bar of silence the woodblock returned either early, late or on time. Participants had to indicate the accuracy of the timing of the woodblock. This novel design required participants to actively maintain the pattern of the woodblock during the silence in order to accurately perform the task. They found that beat related neural oscillations were more pronounced in the silence in musicians than non-musicians, but also that the amplitude at these beat frequencies was positively correlated with performance accuracy (Stupacher et al., 2017).

2.4.4 Brain Stimulation Studies

Finally, one way to explore coordination between regions is to use brain stimulation techniques such as repetitive transcranial magnetic stimulation (rTMS) or transcranial direct current stimulation (tDCS) to investigate the effect it has on performance of imagery tasks. Cathodal tDCS is understood to suppress cortical excitability and therefore have a detrimental effect on performance if the area stimulated is required for the task, whereas anodal tDCS has a facilitation effect, enhancing cortical excitability (Nitsche & Paulus, 2000).

These techniques have been used in a series of pitch and rhythm memory studies by Nora Schaal and colleagues (Schaal et al., 2014; Schaal et al., 2017; Schaal, Williamson, & Banissy, 2013; Schaal, Williamson, et al., 2015). Stimulation was administered at CP3/CP4 locations corresponding to bilateral supramarginal gyrus

(SMG) (BA 40), part of the inferior parietal region between the primary auditory and motor cortices. For non-musicians, anodal stimulation to the left SMG improved pitch recognition and memory recall (Schaal et al., 2013), and pitch memory capacity, but made no difference to rhythm memory capacity (Schaal et al., 2017). Conversely, cathodal stimulation to the left SMG lead to impairments in pitch recognition and recall (Schaal et al., 2014). Interestingly, anodal stimulation to the right SMG led to a facilitation of rhythm memory capacity, but had no effect of pitch memory capacity (Schaal et al., 2017). No effect in pitch memory performance was also found for cathodal stimulation in the right SMG (Schaal et al., 2014). Finally, non-musicians showed performance deficits if rTMS was administered to the left SMG during the retention phase of the task, but not during the encoding phase (Schaal, Williamson, et al., 2015). In contrast, musicians did not show the same deficits in pitch recognition and memory recall after cathodal stimulation to the left SMG, but were worse at pitch recognition after cathodal stimulation to the right SMG (Schaal et al., 2014).

Recently the causal role of the left intraparietal sulcus, a region within the dorsal auditory pathway, in the mental reversal of simple melodies has been investigated using rTMS (Albouy et al., 2017). The task involved a delayed match to sample with two conditions: simple (in which a melody was maintained during the retention period) or manipulation (in which melody had to be reversed in time during the retention period). All melodies were played with an isochronous rhythm. In an initial M/EEG phase prior to stimulation, participant's theta power and phase locking in the left intraparietal sulcus during the retention period for manipulation trials, predicted behavioural performance on the task. Subsequent stimulation of the region boosted performance on manipulation trials, but not the simple maintenance trials. However, this improvement was only seen for rhythmic TMS (at the theta frequency) and not arrhythmic TMS (which provided the

same number of stimulations, but at irregular frequency). The authors conclude that rhythmic TMS entrained theta oscillations in the left IPS and boosted participants' accuracy, thereby establishing that theta activity in the dorsal stream is casually related to memory manipulation (Albouy et al., 2017).

Taken together these results suggest that the left SMG is involved in maintenance of pitch information, specifically in non-musicians, and for higher memory load retention, while the right SMG is involved in maintenance of rhythm memory. While in musicians the right SMG was involved for pitch memory tasks, no comparison of pitch and rhythm memory capacity after stimulation has yet been studied in musicians. Further, for the manipulation of pitch, the left IPS is casually involved in pitch manipulation requiring reversal in time, in a group of participants with a range of musical training from 0 – 17 years. Hence it remains unclear firstly whether in a sample of both musicians and non-musicians there would be hemispheric differences in musical imagery tasks requiring maintenance of pitch and rhythm, and secondly whether differences would also be seen in manipulation of rhythm.

2.5 Individual Differences in Musical Imagery

One determinant of performance accuracy on musical imagery tasks could be subjective imagery vividness. Having participants provide a rating of vividness on a trial-by-trial basis is one way to explore this. Neuroimaging evidence suggests when participants learned pairs of novel songs, (and during the silence between them experienced anticipatory imagery for the upcoming song), more activation in left inferior frontal gyrus and right putamen was seen in participants with increased trial-by-trial vividness scores (Leaver et al., 2009).

Rather than rely on imagery ratings throughout a task, several self-report imagery questionnaires have also been developed to assess imagery vividness (for review see: Hubbard, 2010; Hubbard, 2013). For example, in the Bucknell Auditory Imagery Scale (Halpern, 2015) the vividness subscale requires participants to imagine various sounds such as the sound of a trumpet beginning the piece "Happy Birthday" and rate the vividness of that sound image on a scale of 1 ("no image present at all") to 7 ("as vivid as actual sound"), where 4 corresponds to "fairly vivid". This is similar to the types of vividness surveys previously developed (Hubbard, 2013). Distinctively, the BAIS also has a control subscale, which seeks to measure the ease at which participants can manipulate sound images in their minds. For this second subscale, the same sound images are generated as in the vividness scale. For example, after bringing to mind the trumpet playing "Happy Birthday", the participants are instructed to imagine that the trumpet stops and a violin continues the piece. Participants then rate how easily they could change the sounds on a scale from 1 ("no image present at all") to 7 ("extremely easy to change the image") where 4 would indicate that they "could change the image, but with effort". Hence the two subscales of vividness (BAIS-V) and control (BAIS-C) try to capture the subjective individual differences that may be at work in both maintenance and manipulation imagery paradigms respectively. The scale has been used to screen participants to ensure they are capable of imagery before proceeding with studies (Alonso et al., 2016; Martin et al., 2017).

The BAIS has been shown to correlate with activity and structure of several brain regions. During imagery of familiar songs, participants with higher BAIS-V scores had more cerebral blood flow in right superior temporal gyrus and right dorsolateral prefrontal cortex (Herholz et al., 2012). During a follow up recognition task, which required participants to recognise previously heard or imagined melodies, higher BAIS-V

correlated with activity in the left temporal pole (Herholz & Zatorre, 2012). Activity in the right secondary auditory region was also found to correlate with the total BAIS score during maintenance, and the right intraparietal sulcus during manipulation in a group of trained musicians (Zatorre et al., 2010). Higher BAIS-V has also been found to correlate with grey matter volume in left anterior parietal lobule and left supplementary areas, which are both regions within the general and auditory specific imagery network (Lima et al., 2015).

Finally, the relationship between musical training and imagery vividness has been explored in several studies. The BAIS has been found to correlate moderately ($r \sim 0.3$) with measures of musical training (Halpern, 2015). In addition, evidence from musicians suggests a relationship between subjective vividness and the amount of time spent using imagery in practice (T. Clark & Williamon, 2012). These studies suggest musical training may impact on imagery vividness, but it remains unclear whether imagery vividness is a stable experience or one that can be improved upon with practice.

2.6 Movement and Musical Imagery

Unsurprisingly, given the role of the motor system on musical imagery performance, as discussed above, the impact that movement has on the timing of musical imagery has been well studied. Tapping has been shown to improve temporal prediction (Manning, Harris, & Schutz, 2017; Manning & Schutz, 2013, 2015, 2016; Merchant et al., 2015; Morillon & Baillet, 2017; Morillon et al., 2014) and even in the absence of movement, auditory prediction cues motor preparation (Stephan, 2018).

Given that even short-term motor training enhances the beta-band response in the contralateral sensorimotor region during subsequent passive listening (Ross et al., 2017), it is important to consider the design of musical imagery experiments that incorporate

motor engagement. For example, in the neural entrainment study of Stupacher et al. (2016) discussed above, in which a drum beat pattern was followed by a silent retention period, in a second condition the participants tapped along to the downbeat both during the presentation of the drum pattern and in the retention period. Beat-related steady state evoked potentials (SSEPs) were found in both the non-tapping and tapping condition during the listening period of the drumbeat, but during the silent retention period, SSEPs were only seen in the tapping condition. The two conditions (tapping and non-tapping) were counterbalanced, but with only 14 participants in total, no analysis was done to compare the two groups. It is unclear whether the short-term training of tapping along to the beats for the "tapping first" group, may have influenced subsequent implicit or explicit imagery during the silent retention period, and whether SSEPs may have been stronger for the group whom did their tapping condition first.

Similarly, several temporal prediction studies have compared tapping and "no-movement" conditions (Manning et al., 2017; Manning & Schutz, 2013, 2015, 2016). In these studies participants heard a simple isochronous pattern 3 times, with an accent on every 4th beat (downbeat). After the 4th downbeat the subsequent beats weren't sounded, but the next downbeat came in either slightly early, late or on time (Manning & Schutz, 2013). Though imagery is not mentioned these studies, we argue this task would have required at least implicit if not explicit imagery to perform. The authors found that in the movement condition, the more motor effectors used to do the task (i.e. tapping with drum stick and using arm / wrist, rather than just finger tapping), the better the performance accuracy (Manning et al., 2017). Improvements in performance were not due to the auditory feedback gained through the act of tapping, but through the movement of tapping itself (Manning & Schutz, 2015). Interestingly, those in the non-auditory feedback condition (who heard white noise when the auditory pattern dropped out) appeared to

show better performance in the no-movement task than those who completed the task in silence (Manning & Schutz, 2015). This could suggest that the white noise made participants amplify their internal imagery in order to overcome the distraction and maintain tempo, and hence led to better performance. However, this is speculation, and follow up tests would be required to confirm whether white noise facilitates or impairs imagery performance on such tasks. These studies have also shown that there were no differences between percussionists and non-percussionists in non-movement conditions, suggesting that tapping experience may not be a strong predictor of rhythm imagery ability (Manning & Schutz, 2016).

In summary, the literature suggests a clear improvement after rhythmic movement in temporal prediction, which we argue requires implicit imagery. It remains unclear whether short-term tapping practice shows facilitation in subsequent temporal prediction completed in no-movement contexts and if such facilitation of tapping would also be seen in explicit rhythm imagery.

2.7 Implications for the Real World

Historically, musical imagery ability has been used as a measure of musical aptitude (Gordon, 1989b, 1999; Seashore, 1919), and is said to be well developed in musicians (Aleman et al., 2000). Mozart reportedly experienced his compositions as complete works in his mind through his polyphonic imagery (Agnew, 1922), while Beethoven was (presumably) forced to rely on musical imagery to compose his late symphonies, a period when he was largely or completely deaf (Deutsch & Pierce, 1992). The pianist Glenn Gould had unimpaired hearing but *preferred* to study music by reading it rather than playing it, indicating that musical imagery can be a powerful strategy in and of itself, and not merely a backup strategy necessitated by deafness (Otto, 1990). Among contemporary composers, functional magnetic resonance imaging (fMRI) measurements

show that Sting uses highly similar brain regions when listening to or imagining music (Levitin & Grafton, 2016). Musical imagery can also occur in children, with concert pianist Anna Goldsworthy describing at age nine the first instance of experiencing polyphonic imagery: “I had to sing it in my head, follow its contour, tell its story. Then, when I put the parts together, by a sudden miracle I could hear them all at once. It was as though I had three minds, or three sets of ears, operating in parallel.” (Goldsworthy, 2010, p. 16). Hence the earliest application of the study of musical imagery was limited to music education; teaching young musicians to imagine the desired sound to coordinate their movement to enable that sound to occur (Goldsworthy, 2010). Music education, ensemble playing and even singing ability remain as important applications of musical imagery research today (Greenspon et al., 2017; Keller, 2012; Keller & Appel, 2010; Pfordresher, Halpern, & Greenspon, 2015) however there are many other ways in which this area of research can have benefits.

As we have shown, musical imagery is related to memory, movement, and music perception, hence disorders in any of these domains may be a potential avenue of application for musical imagery. For example, the study of musical imagery and memory disorders such as dementia have been explored (Halpern, Golden, Magdalinou, Witoonpanich, & Warren, 2015), as has musical imagery in motor rehabilitation for Parkinson’s disease or stroke (Lee, Seok, Kim, Park, & Kim, 2018; Schaefer, 2017). The role of musical imagery in auditory hallucinations in clinical and non-clinical populations is also an area of application (Kumar et al., 2014; Linden et al., 2011; Shinosaki et al., 2003). Future research areas could also involve developing musical imagery interventions for those with aided hearing such as recipients of cochlear implants. As yet it is unclear whether hearing loss also leads to a deficit in imagery ability.

Apart from clinical application for rehabilitation, mental imagery for music may provide a window into disorders of consciousness when instructing patients to wilfully change what they are imagining (Monti et al., 2010). In a recent study of 13 healthy adults using fMRI and EEG to decode brain responses during a range of different imagery tasks, the authors found mental arithmetic was the best imagery modality to be classified using machine learning techniques with decodability in fMRI and EEG being significant for 12 and 11 participants respectively. However musical imagery (having the participants imagine their favourite song) was the next best, 10 out of 13 participants showing significant decodability in both imaging modalities (A. H. Harrison, Noseworthy, Reilly, Guan, & Connolly, 2017).

Finally, while it has been shown that running to music results in entrainment of running cadence to the beat (Van Dyck et al., 2015), it is unclear if imagined music would show the same entraining effect. Hence another area of real-world application is the role of musical imagery on sporting activities such as swimming, where access to music listening is not as readily available, or in sporting competitions where the use of music is not allowed.

2.8 Considerations for Future Research

There are several important considerations and directions for future research as a result of this review. First, the various concepts and terminology found in the musical imagery literature were combined into a single conceptual framework, that included explicit and implicit imagery. This overview presented how voluntary musical imagery, (the blue box in Figure 2-1), the focus of this thesis, fits into the broader research literature.

Second, the literature highlights the need for well-designed tasks to induce and test musical imagery. Such tasks should (1) involve objective measures of performance (i.e. accuracy, reaction time); (2) be adaptable and flexible to cater for a range of ability levels; (3) ensure different aspects are studied separately where possible (e.g. pitch and rhythm); (4) confirm that a musical imagery strategy has been used to complete the task; (5) use participants with a range of musical ability to be able to generalise the findings; (6) be designed to clearly distinguish the influence of motor production on imagery performance, if production is being used as a comparison condition. The use of multiple modalities of tools such as questionnaires, behavioural studies and neuroimaging studies is also recommended (Hubbard, 2018).

Third, as auditory-sensorimotor interactions have been the recent focus of implicit imagery tasks such as temporal prediction and beat perception, future research into explicit imagery should seek to explore these interactions, and directly compare imagery and music perception. Specifically in MEG, functional localiser tasks (i.e. passive listening to a series of jittered tones; self-paced index finger tapping from both right and left hands) provide sufficient data to localise bilateral auditory and sensorimotor regions of interest (Morillon & Baillet, 2017; Tal et al., 2017). Even if only right hand tapping data is obtained, a mirror symmetric motor source in the right hemisphere can be used (Ross et al., 2017).

Fourth, the role of oscillations in musical imagery is a growing area of interest with many potential avenues for future research. The beta-band has been argued to be an open line of communication between the auditory and motor regions (Tang et al., 2016), and has been found to be modulated in tasks predicting both timing and pitch in perception (Chang, Bosnyak, & Trainor, 2016; Chang et al., 2018; Fujioka et al., 2012). However, to date most studies in beta-band oscillations and imagery have involved imagined accents

on top of heard tones (Fujioka et al., 2014; Fujioka et al., 2015). When imagery has occurred in silence, the timing of imagined events have been hard to quantify, however the neural entrainment measures using frequency tagging approaches is a potential way to address this (Nozaradan et al., 2011; Stupacher et al., 2016).

Fifth, one way to explore individual differences in musical imagery ability is through measurement of imagery vividness. The BAIS was discussed as a questionnaire that has been used to measure vividness and mental control (Halpern, 2015). To date there have been several behavioural studies that have correlated imagery performance and brain activity and structure, with BAIS scores (Herholz et al., 2012; Lima et al., 2015; Zatorre et al., 2010). However no behavioural tasks have correlated significantly with the control subscale alone, and no correlations between the BAIS and oscillatory activity in the brain have been found.

Lastly, another outstanding question is the role that motor learning plays in explicit imagery performance. With musicians often showing superior imagery ability (Aleman et al., 2000) this may be in part due to the motor training they have undergone, since the brain's response in auditory-motor regions to an instrument's sound is changed through short and long-term practice of that instrument (Krishnan et al., 2018; Pantev & Herholz, 2011; Ross et al., 2017). Hence another avenue for future research is whether production impacts on imagery ability, particularly on rhythm imagery ability.

2.9 References

Agnew, M. (1922). The auditory imagery of great composers. *Psychological Monographs: General and Applied, 31*(1), 279-287.

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron, 94*(1), 193-206. doi:10.1016/j.neuron.2017.03.015

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*(12), 1664-1668. doi:10.1016/S0028-3932(00)00079-8

Alonso, I., Davachi, L., Valabrègue, R., Lambrecq, V., Dupont, S., & Samson, S. (2016). Neural correlates of binding lyrics and melodies for the encoding of new songs. *NeuroImage, 127*, 333-345. doi:10.1016/j.neuroimage.2015.12.018

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390-398. doi:10.1016/j.tics.2012.05.003

Backer, K. C., Binns, M. A., & Alain, C. (2015). Neural dynamics underlying attentional orienting to auditory representations in short-term memory. *The Journal of Neuroscience, 35*(3), 1307-1318. doi:10.1523/jneurosci.1487-14.2015

Bailes, F., Bishop, L., Stevens, C. J., & Dean, R. T. (2012). Mental imagery for musicial changes in loudness. *Frontiers in Psychology, 3*, 525. doi:10.3389/fpsyg.2012.00525

Bishop, L., Bailes, F., & Dean, R. T. (2013). Musical imagery and the planning of dynamics and articulation during performance. *Music Perception: An Interdisciplinary Journal, 31*(2), 97-117. doi:10.1525/mp.2013.31.2.97

Bishop, L., Bailes, F., & Dean, R. T. (2014). Performing musical dynamics: How crucial are musical imagery and auditory feedback for expert and novice musicians? *Music Perception: An Interdisciplinary Journal, 32*(1), 51-66. doi:10.1525/mp.2014.32.1.51

Burianová, H., Marstaller, L., Sowman, P., Tesan, G., Rich, A. N., Williams, M., . . . Johnson, B. W. (2013). Multimodal functional imaging of motor imagery using a novel paradigm. *NeuroImage, 71*, 50-58. doi:10.1016/j.neuroimage.2013.01.001

Cebrian, A. N., & Janata, P. (2010). Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Research, 1342*, 39-54. doi:10.1016/j.brainres.2010.04.026

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2016). Unpredicted pitch modulates beta oscillatory power during rhythmic entrainment to a tone sequence. *Frontiers in Psychology, 7*, 327. doi:10.3389/fpsyg.2016.00327

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2018). Beta oscillatory power modulation reflects the predictability of pitch change. *Cortex, 106*. doi:10.1016/j.cortex.2018.06.008

Chapin, H., Zanto, T., Jantzen, K., Kelso, S., Steinberg, F., & Large, E. (2010). Neural responses to complex auditory rhythms: The role of attending. *Frontiers in Psychology, 1*(224). doi:10.3389/fpsyg.2010.00224

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex, 18*(12), 2844-2854. doi:10.1093/cercor/bhn042

Cheyne, D. O. (2013). MEG studies of sensorimotor rhythms: A review. *Experimental Neurology, 245*, 27-39. doi:10.1016/j.expneurol.2012.08.030

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience, 17*, 718. doi:10.1038/nrn.2016.113

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. doi:10.1017/S0140525X12000477

Clark, T., & Williamon, A. (2012). Imagining the music: Methods for assessing musical imagery ability. *Psychology of Music, 40*(4), 471-493. doi:10.1177/0305735611401126

Corballis, M. C. (1997). Mental rotation and the right hemisphere. *Brain and Language, 57*(1), 100-121. doi:10.1006/brln.1997.1835

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Progress in Brain Research* (Vol. 169, pp. 323-338): Elsevier.

Daselaar, S. M., Porat, Y., Huijbers, W., & Pennartz, C. M. A. (2010). Modality-specific and modality-independent components of the human imagery system. *NeuroImage, 52*(2), 677-685. doi:10.1016/j.neuroimage.2010.04.239

de Borst, A. W., & de Gelder, B. (2016). fMRI-based multivariate pattern analyses reveal imagery modality and imagery content specific representations in primary somatosensory, motor and auditory cortices. *Cerebral Cortex, 27*(8), 3994-4009. doi:10.1093/cercor/bhw211

Deutsch, D. (1972). Effect of repetition of standard and of comparison tones on recognition memory for pitch. *Journal of Experimental Psychology, 93*(1), 156-162. doi:10.1037/h0032496

Deutsch, D., & Pierce, J. R. (1992). The climate of auditory imagery and music. In D. Reiseberg (Ed.), *Auditory Imagery* (pp. 237-260). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations — signalling the status quo? *Current Opinion in Neurobiology, 20*(2), 156-165. doi:10.1016/j.conb.2010.02.015

Farrugia, N., Jakubowski, K., Cusack, R., & Stewart, L. (2015). Tunes stuck in your brain: The frequency and affective evaluation of involuntary musical imagery correlate with cortical structure. *Consciousness and Cognition, 35*, 66-77. doi:10.1016/j.concog.2015.04.020

Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage, 75*, 27-35. doi:10.1016/j.neuroimage.2013.02.044

Foster, N. E. V., & Zatorre, R. J. (2010a). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage, 53*(1), 26-36. doi:10.1016/j.neuroimage.2010.06.042

Foster, N. E. V., & Zatorre, R. J. (2010b). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex, 20*(6), 1350-1359. doi:10.1093/cercor/bhp199

Fujioka, T., Fidali, B., & Ross, B. (2014). Neural correlates of intentional switching from ternary to binary meter in a musical hemiola pattern. *Frontiers in Psychology, 5*, 1257. doi:10.3389/fpsyg.2014.01257

Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *The Journal of Neuroscience, 35*(45), 15187-15198. doi:10.1523/jneurosci.2397-15.2015

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *The Journal of Neuroscience, 32*(5), 1791-1802. doi:10.1523/jneurosci.4107-11.2012

Gabriel, D., Wong, T. C., Nicolier, M., Giustiniani, J., Mignot, C., Noiret, N., . . . Vandel, P. (2016). Don't forget the lyrics! Spatiotemporal dynamics of neural mechanisms spontaneously evoked by gaps of silence in familiar and newly learned songs. *Neurobiology of Learning and Memory, 132*, 18-28. doi:10.1016/j.nlm.2016.04.011

Galton, F. (1880). Statistics of mental imagery. *Mind; a Quarterly Review of Psychology and Philosophy, 5*, 301-318.

Godøy, R. I., & Jørgensen, H. (2001). *Musical imagery*. Lisse, the Netherlands: Swets & Zeitlinger Publishers.

Gordon, E. E. (1985). Research studies in audiation: I. *Bulletin of the Council for Research in Music Education, 84*, 34-50.

Gordon, E. E. (1989). *Learning Sequences in Music: Skill, Content and Patterns*. Chicago, USA: G.I.A. Publications.

Gordon, E. E. (1999). All about audiation and music aptitudes. *Music Educators Journal, 86*(2), 41-44. doi:10.2307/3399589

Grahn, J. A., & Brett, M. (2007). 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., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage, 47*(4), 1894-1903. doi:10.1016/j.neuroimage.2009.04.039

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception: An Interdisciplinary Journal, 34*(5), 585-604. doi:10.1525/mp.2017.34.5.585

Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., . . . Jolicoeur, P. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography. *NeuroImage, 94*, 96-106. doi:10.1016/j.neuroimage.2014.03.020

Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory. *Psychophysiology, 48*(11), 1500-1509. doi:10.1111/j.1469-8986.2011.01234.x

Gunji, A., Ishii, R., Chau, W., Kakigi, R., & Pantev, C. (2007). Rhythmic brain activities related to singing in humans. *NeuroImage, 34*(1), 426-434. doi:10.1016/j.neuroimage.2006.07.018

Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 1-27). Hillsdale, N.J.: L. Erlbaum Associates.

Halpern, A. R. (2003). Cerebral substrates of musical imagery. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 217-230). New York, NY: Oxford University Press; US.

Halpern, A. R. (2015). Differences in auditory imagery self-reported predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain, 25*(1), 37-47. doi:10.1037/pmu0000081

Halpern, A. R., Golden, H. L., Magdalinou, N., Witoonpanich, P., & Warren, J. D. (2015). Musical tasks targeting preserved and impaired functions in two dementias. *Annals of the New York Academy of Sciences, 1337*(1), 241-248. doi:10.1111/nyas.12616

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*(7), 697-704. doi:10.1093/cercor/9.7.697

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*(9), 1281-1292. doi:10.1016/j.neuropsychologia.2003.12.017

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories — A processing based perspective on oscillatory subsequent memory effects. *NeuroImage, 85, Part 2*, 648-655. doi:10.1016/j.neuroimage.2013.05.121

Hari, R., & Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. *Trends in Neurosciences, 20*(1), 44-49. doi:10.1016/S0166-2236(96)10065-5

Hari, R., & Salmelin, R. (2012). Magnetoencephalography: From SQUIDs to neuroscience: Neuroimage 20th Anniversary Special Edition. *NeuroImage, 61*(2), 386-396. doi:10.1016/j.neuroimage.2011.11.074

Harrison, A. H., Noseworthy, M. D., Reilly, J. P., Guan, W., & Connolly, J. F. (2017). EEG and fMRI agree: Mental arithmetic is the easiest form of imagery to detect. *Consciousness and Cognition, 48*, 104-116. doi:10.1016/j.concog.2016.10.006

Harrison, P. M. C., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: Descriptive and explanatory approaches. *Journal of New Music Research, 45*(3), 265-280. doi:10.1080/09298215.2016.1197953

Henry, M. J., & Grahn, J. A. (2017). Music, brain, and movement. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition*. Abingdon: Routledge.

Henry, M. J., & Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Perception, 2*(1), 62-86. doi:10.1163/22134468-00002011

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience, 24*(6), 1382-1397. doi:10.1162/jocn_a_00216

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience, 28*(11), 2352-2360. doi:10.1111/j.1460-9568.2008.06515.x

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2009). Imagery mismatch negativity in musicians. *Annals of the New York Academy of Sciences, 1169*(1), 173-177. doi:10.1111/j.1749-6632.2009.04782.x

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron, 76*(3), 486-502. doi:10.1016/j.neuron.2012.10.011

Hubbard, T. L. (2010). Auditory imagery: empirical findings. *Psychological Bulletin, 136*(2), 302-329.

Hubbard, T. L. (2013). Auditory aspects of auditory imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory Imagery* (pp. 51 - 76). S.I.: Springer New York.

Hubbard, T. L. (2018). Some methodological and conceptual considerations in studies of auditory imagery. *Auditory Perception & Cognition*, 1-36. doi:10.1080/25742442.2018.1499001

Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 45-71). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences, 1169*(1), 58-73. doi:10.1111/j.1749-6632.2009.04579.x

Jakubowski, K., Bashir, Z., Farrugia, N., & Stewart, L. (2018). Involuntary and voluntary recall of musical memories: A comparison of temporal accuracy and emotional responses. *Memory & Cognition, 46*(5), 741-756. doi:10.3758/s13421-018-0792-x

Jakubowski, K., Farrugia, N., Halpern, A. R., Sankarpandi, S., & Stewart, L. (2015). The speed of our mental soundtracks: Tracking the tempo of involuntary musical imagery in everyday life. *Memory & Cognition, 43*(8), 1229-1242. doi:10.3758/s13421-015-0531-5

Jakubowski, K., Farrugia, N., & Stewart, L. (2016). Probing imagined tempo for music: Effects of motor engagement and musical experience. *Psychology of Music, 44*(6), 1274-1288. doi:10.1177/0305735615625791

Janata, P. (2001a). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topography, 13*(3), 169-193. doi:10.1023/A:1007803102254

Janata, P. (2001b). Neurophysiological mechanisms underlying auditory image formation in music. In R. I. Godøy & H. Jørgensen (Eds.), *Musical Imagery*. New York: Routledge.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience, 4*, 186. doi:10.3389/fnhum.2010.00186

Joucla, C., Nicolier, M., Giustiniani, J., Brunotte, G., Noiret, N., Monnin, J., . . . Gabriel, D. (2018). Evidence for a neural signature of musical preference during silence. *International Journal of Psychophysiology, 125*, 50-56. doi:10.1016/j.ijpsycho.2018.02.007

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences, 1252*(1), 206-213. doi:10.1111/j.1749-6632.2011.06439.x

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception: An Interdisciplinary Journal, 28*(1), 27-46. doi:10.1525/mp.2010.28.1.27

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics, 14*(1), 90-94. doi:10.3758/BF03198621

Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain*. Cambridge: The MIT Press.

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin, 129*(5), 723-746.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature, 434*(7030), 158.

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., . . . Scott, S. K. (2018). Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play. *Cerebral Cortex, bhy208*. doi:10.1093/cercor/bhy208

Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., & Pantev, C. (2012). Electromagnetic correlates of musical expertise in processing of tone patterns. *PLoS ONE, 7*(1). doi:10.1371/journal.pone.0030171

Kumar, S., Sedley, W., Barnes, G. R., Teki, S., Friston, K. J., & Griffiths, T. D. (2014). A brain basis for musical hallucinations. *Cortex, 52*, 86-97. doi:10.1016/j.cortex.2013.12.002

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119-159. doi:10.1037/0033-295X.106.1.119

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *The Journal of Neuroscience, 29*(8), 2477-2485. doi:10.1523/jneurosci.4921-08.2009

Lee, S. Y., Seok, H., Kim, S.-H., Park, M., & Kim, J. (2018). Immediate effects of mental singing while walking on gait disturbance in hemiplegic stroke patients: A feasibility study. *Annals of Rehabilitation Medicine, 42*(1), 1-7. doi:10.5535/arm.2018.42.1.1

Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., . . . Jolicœur, P. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia, 51*(13), 2939-2952. doi:10.1016/j.neuropsychologia.2013.08.003

Lehtelä, L., Salmelin, R., & Hari, R. (1997). Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex. *Neuroscience Letters, 222*(2), 111-114. doi:10.1016/S0304-3940(97)13361-4

Levitin, D. J., & Grafton, S. T. (2016). Measuring the representational space of music with fMRI: a case study with Sting. *Neurocase, 22*(6), 548-557. doi:10.1080/13554794.2016.1216572

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences, 39*(8), 527-542. doi:10.1016/j.tins.2016.06.003

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., . . . Scott, S. K. (2015). Feel the noise: relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex, 25*(11), 4638-4650. doi:10.1093/cercor/bhv134

Linden, D. E. J., Thornton, K., Kuswanto, C. N., Johnston, S. J., van de Ven, V., & Jackson, M. C. (2011). The brain's voices: Comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex, 21*(2), 330-337. doi:10.1093/cercor/bhq097

Lotze, M., Scheler, G., Tan, H.-R. M., Braun, C., & Birbaumer, N. (2003). The musician's brain: functional imaging of amateurs and professionals during performance and imagery. *NeuroImage, 20*(3), 1817-1829. doi:10.1016/j.neuroimage.2003.07.018

Loui, P., Wessel, D. L., & Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception: An Interdisciplinary Journal, 27*(5), 377-388. doi:10.1525/mp.2010.27.5.377

MacKisack, M., Aldworth, S., Macpherson, F., Onians, J., Winlove, C., & Zeman, A. (2016). On picturing a candle: The prehistory of imagery science. *Frontiers in Psychology, 7*, 515. doi:10.3389/fpsyg.2016.00515

Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in Psychology, 4*, 1008. doi:10.3389/fpsyg.2013.01008

Mandikal Vasuki, P. R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research, 342*, 112-123. doi:10.1016/j.heares.2016.10.008

Manning, F. C., Harris, J., & Schutz, M. (2017). Temporal prediction abilities are mediated by motor effector and rhythmic expertise. *Experimental Brain Research, 235*(3), 861-871. doi:10.1007/s00221-016-4845-8

Manning, F. C., & Schutz, M. (2013). "Moving to the beat" improves timing perception. *Psychonomic bulletin & review, 20*(6), 1133-1139. doi:10.3758/s13423-013-0439-7

Manning, F. C., & Schutz, M. (2015). Movement enhances perceived timing in the absence of auditory feedback. *Timing & Time Perception, 3*(1-2), 3-12. doi:10.1163/22134468-03002037

Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research, 80*(4), 532-542. doi:10.1007/s00426-015-0678-5

Martin, S., Mikutta, C., Leonard, M. K., Hungate, D., Koelsch, S., Shamma, S., . . . Pasley, B. N. (2017). Neural encoding of auditory features during music perception and imagery. *Cerebral Cortex*, 1-12. doi:10.1093/cercor/bhx277

McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience, 6*, 285. doi:10.3389/fnhum.2012.00285

Merchant, H., Grahn, J. A., Trainor, L. J., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664). doi:10.1098/rstb.2014.0093

Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restorative Neurology And Neuroscience, 25*(3-4), 411-431.

Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., . . . Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. *The New England journal of medicine, 362*(7), 579-589. doi:10.1056/NEJMoa0905370

Moore, M. E. (2011). Imagination and the mind's ear. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 71*(11-A), 4046.

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences, 114*(42), E8913-E8921. doi:10.1073/pnas.1705373114

Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications, 5*, 5255. doi:10.1038/ncomms6255

Mullensiefen, D., Ginras, B., Musil, J. J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE, 9*(2), e89642. doi:10.1371/journal.pone.0089642

Müller, N., Keil, J., Obleser, J., Schulz, H., Grunwald, T., Bernays, R.-L., . . . Weisz, N. (2013). You can't stop the music: Reduced auditory alpha power and coupling between auditory and memory regions facilitate the illusory perception of music during noise. *NeuroImage, 79*, 383-393. doi:10.1016/j.neuroimage.2013.05.001

Neisser, U. (1972). Changing conceptions of imagery. In P. W. Sheehan (Ed.), *The functions and nature of imagery* (pp. 233-251). New York: Academic Press.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology, 527*(3), 633-639. doi:10.1111/j.1469-7793.2000.t01-1-00633.x

Nolden, S., Grimault, S., Guimond, S., Lefebvre, C., Bermudez, P., & Jolicoeur, P. (2013). The retention of simultaneous tones in auditory short-term memory: A magnetoencephalography study. *NeuroImage, 82*, 384-392. doi:10.1016/j.neuroimage.2013.06.002

Nozaradan, S., Peretz, I., & Keller, P. E. (2016). Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Scientific Reports, 6*, 20612. doi:10.1038/srep20612

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience, 31*(28), 10234-10240. doi:10.1523/jneurosci.0411-11.2011

Otsuka, A., Tamaki, Y., & Kuriki, S. (2008). Neuromagnetic responses in silence after musical chord sequences. *NeuroReport, 19*(16), 1637-1641. doi:10.1097/WNR.0b013e32831576fa

Otto, F. (1990). *Glenn Gould: a life and variations*. New York: Vintage Books.

Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience & Biobehavioral Reviews, 35*(10), 2140-2154. doi:10.1016/j.neubiorev.2011.06.010

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the action simulation for auditory prediction (ASAP) hypothesis. *Frontiers in systems neuroscience, 8*, 57. doi:10.3389/fnsys.2014.00057

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50*(1), 302-313. doi:10.1016/j.neuroimage.2009.12.019

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences, 112*(33), 10089-10092. doi:10.1073/pnas.1504933112

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences, 19*(10), 590-602. doi:10.1016/j.tics.2015.08.003

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The multi-modal imagery association (MMIA) model. *Music Perception: An Interdisciplinary Journal, 32*(3), 242-253. doi:10.1525/mp.2015.32.3.242

Pfurtscheller, G., & Lopes da Silva, F. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology, 110*(11), 1842-1857. doi:10.1016/S1388-2457(99)00141-8

Puschmann, S. (2013). Pitch-induced responses in the right auditory cortex correlate with musical ability in normal listeners. *NeuroReport, 24*(15), 841-845. doi:10.1097/WNR.0b013e3283650e1e

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*(1), 1-24. doi:10.1037/h0034650

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*(1), 16-45. doi:10.1037/0033-295X.88.1.16

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research, 271*(1), 16-25. doi:10.1016/j.heares.2010.09.001

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*(6), 718-724. doi:10.1038/nn.2331

Reisberg, D. (1992). *Auditory imagery* (D. Reisberg Ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Ross, B., Barat, M., & Fujioka, T. (2017). Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced beta-band oscillations during perception. *The Journal of Neuroscience, 37*(24), 5948-5959. doi:10.1523/jneurosci.3613-16.2017

Schaal, N. K., Banissy, M. J., & Lange, J. (2015). The rhythm span task: Comparing memory capacity for musical rhythms in musicians and non-musicians. *Journal of New Music Research, 44*(1), 3-10. doi:10.1080/09298215.2014.937724

Schaal, N. K., Krause, V., Lange, K., Banissy, M. J., Williamson, V. J., & Pollok, B. (2014). Pitch memory in nonmusicians and musicians: Revealing functional differences using transcranial direct current stimulation. *Cerebral Cortex*. doi:10.1093/cercor/bhu075

Schaal, N. K., Pollok, B., & Banissy, M. J. (2017). Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Scientific Reports, 7*, 42456. doi:10.1038/srep42456

Schaal, N. K., Williamson, V. J., & Banissy, M. J. (2013). Anodal transcranial direct current stimulation over the supramarginal gyrus facilitates pitch memory. *European Journal of Neuroscience, 38*(10), 3513-3518. doi:doi:10.1111/ejn.12344

Schaal, N. K., Williamson, V. J., Kelly, M., Muggleton, N. G., Pollok, B., Krause, V., & Banissy, M. J. (2015). A causal involvement of the left supramarginal gyrus

during the retention of musical pitches. *Cortex, 64*, 310-317. doi:10.1016/j.cortex.2014.11.011

Schaefer, R. S. (2014). Mental representations in musical processing and their role in action-perception loops. *Empirical musicology review : EMR, 9*(3-4), 161-176.

Schaefer, R. S. (2017). Music in the brain: imagery and memory. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition*. Abingdon: Routledge.

Schaefer, R. S., Desain, P., & Suppes, P. (2009). Structural decomposition of EEG signatures of melodic processing. *Biological Psychology, 82*(3), 253-259. doi:10.1016/j.biopsycho.2009.08.004

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011a). Decomposing rhythm processing: Electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological Research, 75*(2), 95-106.

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011b). Music perception and imagery in EEG: alpha band effects of task and stimulus. *International Journal of Psychophysiology, 82*(3), 254-259. doi:10.1016/j.ijpsycho.2011.09.007

Schleuter, S. L. (1993). The relationship of AMMA scores to sightsinging, dictation, and SAT scores of university music majors. *Contributions to Music Education, 20*, 57-63.

Seashore, C. E. (1919). Musical imagery and imagination *The Psychology of Musical Talent*. New York, NY, US: Silver, Burdett & Company. .

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701-703. doi:10.1126/science.171.3972.701

Shinosaki, K., Yamamoto, M., Ukai, S., Kawaguchi, S., Ogawa, A., Ishii, R., . . . Takeda, M. (2003). Desynchronization in the right auditory cortex during musical hallucinations: A MEG study. *Psychogeriatrics, 3*(2), 88-92. doi:10.1046/j.1479-8301.2003.00009.x

Spitzer, B., & Haegens, S. (2017). Beyond the status quo: A role for beta oscillations in endogenous content (re)activation. *eneuro, 4*(4). doi:10.1523/eneuro.0170-17.2017

Strawson, P. F. (1974). Imagination and perception *Freedom and resentment, and other essays* (pp. 50–72). London: Methuen.

Stupacher, J., Witte, M., Hove, M. J., & Wood, G. (2016). Neural entrainment in drum rhythms with silent breaks: Evidence from steady-state evoked and event-related potentials. *Journal of Cognitive Neuroscience, 28*(12), 1865-1877. doi:10.1162/jocn_a_01013

Stupacher, J., Wood, G., & Witte, M. (2017). Neural entrainment to polyrhythms: A comparison of musicians and non-musicians. *Frontiers in Neuroscience, 11*(208). doi:10.3389/fnins.2017.00208

Sutherland, M. E., Paus, T., & Zatorre, R. J. (2013). Neuroanatomical correlates of musical transposition in adolescents: A longitudinal approach. *Frontiers in systems neuroscience, 7*. doi:10.3389/fnsys.2013.00113

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017). Neural entrainment to the beat: the "missing-pulse" phenomenon. *The Journal of Neuroscience, 37*(26), 6331-6341. doi:10.1523/jneurosci.2500-16.2017

Tang, H., Crain, S., & Johnson, B. W. (2016). Dual temporal encoding mechanisms in human auditory cortex: Evidence from MEG and EEG. *NeuroImage, 128*, 32-43. doi:10.1016/j.neuroimage.2015.12.053

Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2017). Effects of sad and happy music on mind-wandering and the default mode network. *Scientific Reports, 7*(1), 14396. doi:10.1038/s41598-017-14849-0

Teki, S., & Kononowicz, T. W. (2016). Commentary: Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Frontiers in Neuroscience, 10*(389). doi:10.3389/fnins.2016.00389

Thut, G., Schyns, P., & Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Frontiers in Psychology, 2*(170). doi:10.3389/fpsyg.2011.00170

van Dijk, H., Nieuwenhuis, I. L. C., & Jensen, O. (2010). Left temporal alpha band activity increases during working memory retention of pitches. *European*

Journal of Neuroscience, 31(9), 1701-1707. doi:10.1111/j.1460-9568.2010.07227.x

Van Dyck, E., Moens, B., Buhmann, J., Demey, M., Coorevits, E., Dalla Bella, S., & Leman, M. (2015). Spontaneous entrainment of running cadence to music tempo. *Sports Medicine - Open, 1*(1), 15. doi:10.1186/s40798-015-0025-9

Vlek, R. J., Schaefer, R. S., Gielen, C. C. A. M., Farquhar, J., & Desain, P. (2011). Shared mechanisms in perception and imagery of auditory accents. *Clinical Neurophysiology, 122*(8), 1526-1532.

Weir, G., Williamson, V. J., & Mullensiefen, D. (2015). Increased involuntary musical mental activity is not associated with more accurate voluntary musical imagery. *Psychomusicology: Music, Mind & Brain, 25*(1), 48-57. doi:10.1037/pmu0000076

Weisz, N., Hartmann, T., Müller, N., & Obleser, J. (2011). Alpha rhythms in audition: Cognitive and clinical perspectives. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00073

Williamson, V. J., Jilka, S. R., Fry, J., Finkel, S., Müllensiefen, D., & Stewart, L. (2012). How do “earworms” start? Classifying the everyday circumstances of Involuntary Musical Imagery. *Psychology of Music, 40*(3), 259-284. doi:10.1177/0305735611418553

Yoo, S.-S., Lee, C. U., & Choi, B. G. C. A. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport, 12*(14), 3045-3049.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547-558.

Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia, 31*(3), 221-232. doi:10.1016/0028-3932(93)90086-F

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron, 47*(1), 9-12. doi:10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22*(4), 775-789. doi:10.1162/jocn.2009.21239

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*(1), 29-46. doi:10.1162/jocn.1996.8.1.29

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience, 37*(9), 1421-1434. doi:10.1111/ejn.12140

Chapter 3

The Pitch Imagery Arrow Task:

Effects of Musical Training, Vividness, and Mental Control

Rebecca W. **Gelding**[a,c], William F. **Thompson**[b,c], Blake W. **Johnson**[a,c]

[a] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[b] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

[c] ARC Centre of Excellence in Cognition and its Disorders

This chapter has been published as:

Gelding, R.W., Thompson, W.F. & Johnson, B.W. (2015). The Pitch Imagery Arrow Task: Effects of musical training, vividness, and mental control, *PLoS ONE 10*(3): e0121809. doi:10.1371/journal.pone.0121809

Author Contributions

All authors conceived and designed the experiment. R.W.G. conducted the experiment. R.W.G. and B.W.J. analysed the data. All authors discussed the results, wrote and edited the manuscript.

Abstract

Musical imagery is a relatively unexplored area, partly because of deficiencies in existing experimental paradigms, which are often difficult, unreliable, or do not provide objective measures of performance. Here we describe a novel protocol, the Pitch Imagery Arrow Task (PIAT), which induces and trains pitch imagery in both musicians and non-musicians. Given a tonal context and an initial pitch sequence, arrows are displayed to elicit a scale-step sequence of imagined pitches, and participants indicate whether the final imagined tone matches an audible probe. It is a staircase design that accommodates individual differences in musical experience and imagery ability. This new protocol is used to investigate the roles that musical expertise, self-reported auditory imagery vividness and mental control play in imagery performance. Performance on the task was significantly better for participants who employed a musical imagery strategy compared to participants who used an alternative cognitive strategy and positively correlated with scores on the Control subscale from the Bucknell Auditory Imagery Scale (BAIS). Multiple regression analysis revealed that Imagery performance accuracy was best predicted by a combination of strategy use and scores on the Vividness subscale of BAIS. These results confirm that competent performance on the PIAT requires active musical imagery and is very difficult to achieve using alternative cognitive strategies. Auditory imagery vividness and mental control were more important than musical experience in the ability to perform manipulation of pitch imagery.

3.1 Introduction

Musical imagery can be described as "hearing a tune in your head" (Zatorre & Halpern, 2005, p. 9). It is a common, everyday experience even for those with no musical training. Despite the ubiquity of musical imagery, research on the topic has often examined musicians and non-musicians separately. This is largely because of deficiencies in existing experimental protocols for inducing and measuring musical imagery. In some cases, imagery tasks are too difficult for non-musicians to complete (Zatorre et al., 2010); in other cases, they are too easy for musicians and susceptible to ceiling effects for this population (Yumoto et al., 2005). Other studies have considered musicians with a range of abilities and correlated their performance on imagery tasks with years of musical training (Bailes et al., 2012), or cumulative hours of training (Foster et al., 2013). In general, musicians perform better than non-musicians on musical imagery tasks (Hubbard, 2010; Zatorre et al., 2010). However, it is not always clear how to interpret such findings because musical knowledge is sometimes needed in order to understand and/or complete these tasks.

Most studies of musical imagery require participants to maintain an image of familiar melodies or scales (Zatorre, 2012). These tasks require a variety of judgments including: pitch judgments about two syllables of the lyrics associated with an imagined tune (Zatorre et al., 1996); comparing the similarity of two imagined instrumental timbres (Halpern et al., 2004); imagining the continuation of ascending musical scales (Cebrian & Janata, 2010; Janata & Paroo, 2006); imagining familiar music during gaps in the presentation (Kraemer et al., 2005; Schaefer et al., 2011b); imagining the continuation of a familiar melody and judging an audible tone for accuracy, as the last tone of the melody (Herholz et al., 2008).

Other experimental paradigms are designed to induce dynamic musical imagery, and hence require participants to shift from one musical image to another through effortful manipulation. These paradigms have included tasks that require mentally transposing a melody into a different key or imaging the pitches of a melody in reverse (retrograde) order (Foster et al., 2013; Zatorre et al., 2010).

This distinction – between maintenance and manipulation of mental imagery – has also been the subject of investigation in other modalities, particularly in visual imagery. Studies have shown that the ability to form vivid visual images is psychometrically distinct from the ability to manipulate those images in space. In one study, visual artists were found to excel at maintaining object imagery but had difficulty with spatial manipulation of images, whereas engineers and scientists exhibited the opposite pattern of performance (Kozhevnikov, Kosslyn, & Shephard, 2005). In the auditory domain, Hansen, Wallentin, and Vuust (2013) found that musicians could recall longer sequences of spoken digits than non-musicians, but they were no better than non-musicians at a backward digit span task that required mental manipulation of that verbal image. These findings suggest that tasks requiring a large store of musical sequences in short-term memory, such as those required in maintenance paradigms, are biased towards musicians. As such, tasks requiring the mental manipulation of musical materials with minimal burden on short-term memory may be better suited to study the role of training on musical imagery.

Another area of individual differences that has received relatively little attention is the vividness of auditory imagery among nonclinical populations (Hubbard, 2013). The Bucknell Auditory Imagery Scale (BAIS) is a self-report questionnaire that incorporates a *vividness* subscale (BAIS-V) and a *control* subscale (BAIS-C), with the former measuring the subjective clarity of an image and the latter measuring the ease with which

participants can change or manipulate an auditory image at will (Halpern, 2015). Previous studies have shown that results from the BAIS, particularly BAIS-V, correlate with performance on musical imagery tasks (Herholz et al., 2012; Pfordresher & Halpern, 2013; Zatorre et al., 2010). The BAIS-C has also been shown to correlate significantly with performance on a pitch discrimination task where participants were required to indicate which of two tones was higher in pitch (Pfordresher & Halpern, 2013). Interestingly, BAIS scores have repeatedly been shown to correlate only mildly ($r \sim .30$) with musical training or experience (Herholz et al., 2012; Pfordresher & Halpern, 2013).

This investigation employed a novel protocol for evaluating musical imagery – the *Pitch Imagery Arrow Task* (PIAT). Several considerations informed the design of this task. First, a number of paradigms that purport to measure musical imagery do not provide any objective behavioural confirmation that participants have actually used musical imagery (Hubbard, 2010; Zatorre & Halpern, 2005). For example, Kraemer et al. (2005) had participants passively listen to familiar and unfamiliar music that had silent pauses of 2–5 s inserted. They then examined brain activity during the period of silence. Although subjects were not instructed to imagine the tunes, all participants reported musical imagery during gaps in the familiar music but not during gaps in the unfamiliar music. Yoo et al. (2001) had participants familiarise themselves with a single pitch outside of the scanner and then were required to imagine the same pitch while their brain was scanned using fMRI. In these studies, imagery was assumed or argued to have occurred either because participants were explicitly instructed to form images or because the investigators created a context in which imagery was likely to occur (Hubbard, 2010; Zatorre & Halpern, 2005).

Second, although some imagery tasks have an objective behavioural outcome measure, they are so simple or repetitive that musical imagery may be unnecessary to

perform the task. For example, Janata and Paroo used familiar ascending scales in a number of their studies, in order to "force listeners to move their mental images in pitch space" (Janata & Paroo, 2006, p. 836). However, these authors acknowledged that by using familiar scales and confirming their use of imagery only for the last note in the sequence (i.e. the tonic), participants may have used the initial (tonic) note to infer the final note of the scale (tonic one octave above the initial scale note). That is, they were not obliged to imagine each note of the scale (Janata & Paroo, 2006). Herholz et al. (2008) required participants to listen to one of nine familiar nursery rhyme introductions (16 repetitions per melody), then imagine the series of missing tones and were tested on the last note of the sequence. This repeated exposure may have led to a learned association between the start of the sequences and the sounded last note, rather than the use of musical imagery to continue the missing tones. To combat this they asked participants whether they had used "any other strategy than imagining the melody, in order to fulfil the task of judging the correctness of the test tone" (Herholz et al., 2008, p. 2354). Having participants confirm the type of strategy used to complete a given imagery task can be an additional way of ensuring auditory imagery is used rather than an alternative cognitive strategy (Williamson, Baddeley, & Hitch, 2010), though such a question should be worded in a way that does not bias participants in their response.

At the other extreme are protocols that are so difficult that only expert musicians can perform them, such as the mental reversal of familiar melodies (Zatorre, et al., 2010; see also: Foster et al., 2013; Zatorre, 2012). As only highly trained participants can complete this task, this limits the range and generalisability of the results.

Finally it has been said that the study of imagery stands precisely at the intersection of two key branches of cognitive psychology - perception and memory (Neisser, 1972). As such, we have included both a Perception control condition and a

Mental Arithmetic control condition to be able to compare imagery performance (accuracy and reaction time) with both music perception and non-musical working memory.

The PIAT has a number of advantages over existing protocols for evaluating imagery. Specifically, the task (1) requires a behavioural response to objectively measure accuracy and response times of imagery performance; (2) is extremely difficult to successfully perform using alternative cognitive strategies other than pitch imagery; (3) employs novel rather than familiar sequences of pitches that cannot be anticipated in advance; (4) employs a range of difficulties implemented in a staircase design, such that it can induce imagery in participants with a wide range of musical experience; (5) incorporates a pitch perception control condition, permitting the assessment of differences in strategies and brain mechanism for imagery versus perception; (6) incorporates a mental arithmetic control condition, permitting the assessment of differences in strategies and brain mechanisms for imagery tasks that employ different cognitive computations.

Our primary goal was to verify the efficacy of the PIAT in inducing and training pitch imagery in musicians and non-musicians with a wide range of musical experience. We also investigated the role of musical training, imagery vividness, and mental control in predicting performance. We hypothesised that (1) participants who used a pitch imagery strategy would show significantly better performance than participants who employed alternative cognitive strategies; (2) successful task performance should be determined more by an individual's vividness and control of musical images (as indexed by the BAIS), than one's history of musical training.

3.2 Materials and Methods

3.2.1 Participants

24 trained musicians and 16 non-musicians were recruited for the study. All participants self-reported to being right-handed, having normal or corrected-to-normal vision and normal hearing. Only right-handed individuals were recruited. Participants were classified as musicians if they had more than 5 years of continuous formal music lessons and have been actively playing their instrument in the last 2 years. Non-musicians were defined as those with less than 2 years formal training, or those who had been non-active in their instrument for more than 10 years. All participants were classified as either a musician or a non-musician based on these criteria. All participants provided written consent and all procedures were approved by the Macquarie University Human Research Ethics Committee. Table 3-1 summarises the characteristics of the two groups. The musicians and non-musicians did not differ significantly in age, gender, daily hours spent listening to music, or education, but they did differ significantly in the Musical Experience Index (MEI). This was calculated as the number of years spent actively playing an instrument / singing, either through formal lessons or self-taught, divided by current age to obtain a percentage of musical experience over the lifetime. For example, if a participant had piano lessons for 2 years, then stopped playing, and later taught themselves to play guitar for 3 years, and are currently aged 25, their musical experience index was (2 + 3)/25 = 0.2. However, if these lessons had happened concurrently then the total years of playing music would be 3, and so their MEI would be 3/25 = .12. This index was used to normalise the musical training across the wide age range of participants. The musicians on average had spent 45% of their life's years actively participating in music, while non-musicians had spent on average 12%, as seen in Table 3-1.

Table 3-1

Summary of the demographic details of the participants.

	Mean Age (Range)	**Number of Females**	**Mean MEI (range)**	**Mean daily hours spent listening to music(range)**
Musician (N=24)	26.2 (18-48)	15	.45 (.16-.72)	2.5 (0.12-10)
Non-Musician (N=16)	22.6 (18-41)	7	.12 (0-.28)	1.5 (0.25-4)
Total Sample (N=40)	24.7 (18-48)	22	.34 (0-.72)	2.1(0.12-10)

3.2.2 Stimuli

3.2.2.1 Pitch Imagery Condition

The PIAT, as outlined in Figure 3-1, consists of three successive components: (1) a *set-up* component in which the participant listens to a starting sequence of computer-generated piano notes. Each successive note is immediately up or down the major scale relative to the preceding note, with the scale direction of the next note (ascending or descending) being random and unpredictable from note to note. (2) An *imagery* (continuation) component in which the piano notes stop while the participant is prompted to imagine a series of 1-5 notes continuing from the last heard note, prompted by vertical up or down arrows which indicate the scale direction for each succeeding note to be imagined; (3) A *probe* component which presents a heard piano note that either matches or does not match the last of the notes in the imagery sequence. After hearing the probe participants were required to indicate with a button press if the probe is a match or a mismatch to the last note of the imagined sequence.

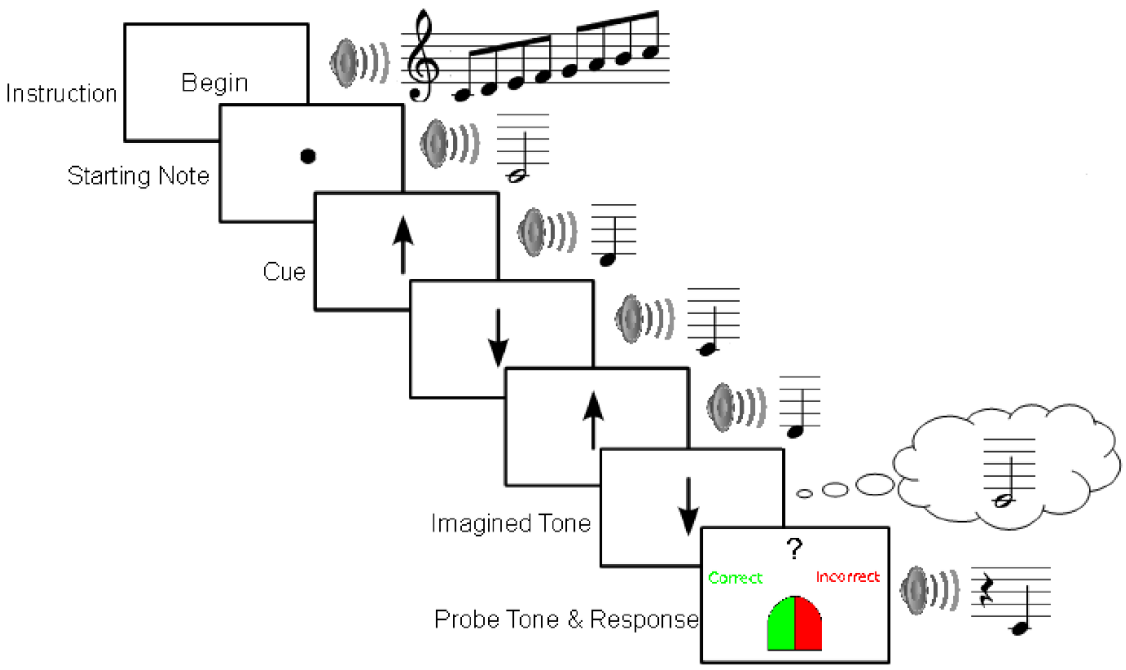

Figure 3-1: Schematic of the Pitch Imagery Arrow Task (PIAT)

The set-up component of each trial began with an ascending major scale that started and ended on the tonic, to provide participants with a tonal context that they could use as a frame of reference for their subsequent judgments (Halpern, 1992). Following the key-defining scale, the starting note of the test sequence is played for a duration of 2 s, and a fixation circle appears in the middle of the computer screen. Following the starting note, each successive note was randomly selected from a position immediately up or down the scale from the last note at a rate of one per second and played for a duration of 500 ms. Each note was accompanied on the visual display with an up or down arrow that indicated the scale direction from the preceding note.

In the imagery component, one to five arrows were then presented in silence at the same rate as the preceding note / arrow combinations of the set-up phase and participants were required to imagine the next scale note up or down from the previous (heard or imagined) note as indicated by the scale direction arrows.

In the probe component of each trial, a target screen with an image of a mouse indicating left click for “Correct” and right click for “Incorrect” was displayed for 1 second to alert participants to an impending probe. A probe tone was then played, and participants were required to indicate whether the probe tone matched or did not match the final note of the imagery component. If no response was recorded within 4 seconds the trial was coded as a missed trial, and participants received a warning message to respond more quickly on future trials. Feedback was provided on each trial to advise the participant whether their response was accurate or not.

The PIAT was programmed with five levels of imagery difficulty, corresponding to the number of sequential notes (from 1-5) in the imagery component. Each imagery level contained three stages that manipulated the complexity of the set-up component, in terms of the maximum number of audible notes in the set-up sequence, the key signature for the sequence, and whether the starting note was a tonic or dominant. In particular, in stages 2 and 3 the key signatures were randomised from a possible 5 key signatures (C Major, C# Major, D Major, Eb Major or E Major). These were set so as not to repeat the previous trial to ensure variability in each stage (Janata & Paroo, 2006). Table 3-2 summarises the attributes for each level and stage. The number of sequential notes in the set-up component was also randomised between trials so that participants were unable to predict when they would be required to begin the imagery component.

The probe tone was a match for 50% of the trials. For incorrect probe trials the probe tone was always in the same key and within 2 steps of the correct answer. The maximum possible range of notes for each trial was set to 3 scale steps up or down from the starting note. For example, for a trial in C Major, beginning on Middle C (C4), the tones (both heard or to be imagined) were within the range of G below middle C (G3) and F above middle C (F4).

Table 3-2

Summary of Level and Stage Design of the Imagery and Mental Arithmetic (MA) conditions.

Level	Stage	Key Signature	Starting Note	No. of arrow with heard tones	No. of imagined tones (Imagery)	Starting Number for MA	No. of arrows for MA	Numeral presented with arrow for MA
1	1	C	tonic	3	1	1	4	1
	2	*C, C#, D, Eb, E	tonic	*3 – 4		*1 – 5	*4 – 5	
	3	*C, C#, D, Eb, E	dominant	*3 – 6		*1 – 5	*4 – 7	
2	1	C	tonic	*3 – 4	2	1	*5 – 6	*1 – 2
	2	*C, C#, D, Eb, E	tonic	*3 – 5		*1 – 5	*5 – 7	
	3	*C, C#, D, Eb, E	dominant	*3 – 6		*1 – 5	*5 – 8	
3	1	C	tonic	*3 – 4	3	1	*6 – 7	*1 – 3
	2	*C, C#, D, Eb, E	tonic	*3 – 5		*1 – 5	*6 – 8	
	3	*C, C#, D, Eb, E	dominant	*3 – 6		*1 – 5	*6 – 9	
4	1	C	tonic	*3 – 4	4	1	*7 – 8	*1 – 4
	2	*C, C#, D, Eb, E	tonic	*3 – 5		*1 – 5	*7 – 9	
	3	*C, C#, D, Eb, E	dominant	*3 – 6		*1 – 5	*7 – 10	
5	1	*C, C#, D, Eb, E	tonic	*3 – 4	5	1	*8 – 9	*1 – 5
	2	*C, C#, D, Eb, E	dominant	*3 – 5		*1 – 5	*8 – 10	
	3	*C, C#, D, Eb, E	*tonic or dominant	*3 – 6		*1 – 5	*8 – 11	

The Perception condition corresponded to the first five columns of the table. *denotes when a random variable from those listed could be used at that stage. Only the key signature was set so as not to repeat for a concurrent trial

3.2.2.2 Pitch Perception Condition

The Pitch Perception trials were identical to the imagery trials, except that arrows were always paired with heard notes during the continuation component. In this case the last note in the sequence then became the correct response for the probe.

3.2.2.3 Mental Arithmetic Condition

The Mental Arithmetic trial's start screen showed "Begin Mental Arithmetic", then a number appeared on the screen which was the starting point of the sequence. The up and down arrows also included a number at their point which indicated how much to increase (up) or decrease (down) the running total by. After a random number of arrows were presented, which varied from a minimum of 4 (Level 1, Stage 1) to a maximum of 11 (Level 5, Stage 3), the same target screen with the mouse image then appeared. After 1 second, a number appeared at the top of the screen indicating the probe number. Participants then responded in the same way as the music trials to confirm if the number was correct or incorrect. Incorrect answers were presented on 50% of the trials and were always 1 integer away from the correct number. The mental arithmetic trials also increased in difficulty as the levels progressed, as per Table 3-2, however the sequences were set to remain in a range of 0 – 10. This range was determined after pilot testing as participants reported being unable to successfully and consistently compute larger numbers at the required rate of one arrow per second (as per the music trials).

3.2.2.4 Bucknell Auditory Imagery Scale

In order to assess the role of self-reported auditory imagery vividness and control on pitch imagery performance, the participants also completed the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015). The scale includes two subscales, for vividness (BAIS-V) and control (BAIS-C), both of which have 14 items each. BAIS-V questions required participants to rate how clearly they could imagine a particular auditory image

from 1 – 7, such as a trumpet beginning to play “Happy Birthday”, with 1 indicating that no image was present at all, 4 being fairly vivid and 7 being as vivid as actual sound. BAIS-C questions required participants to rate similarly from 1 – 7, how easily they could then change an image from, for example, the trumpet beginning the piece to a violin continuing the song. Previous studies have shown that results from the BAIS-V, correlate with performance on musical imagery tasks (Herholz et al., 2012; Pfordresher & Halpern, 2013; Zatorre et al., 2010), but no previous studies have reported a correlation with BAIS-C and imagery performance.

3.2.3 Procedure

Presentation® software (www.neurobs.com) was used to control the experiment and to record responses. Acoustic stimuli were generated from the 'Piano' instrument sound by Finale 2012 software (Makemusic Inc; Eden Prairie, MN) and exported as .wav files for use in Presentation®.

Upon being seated in front of the computer, participants were given a sound check, whereby they could manually adjust the volume of the tones to a suitable level. They were then introduced to the three types of trials, and given a simple example of a Pitch Imagery trial and a Mental Arithmetic trial. Participants were informed that no movement or humming was allowed to assist them with the task, but they should “as vividly as possible, imagine the tones and keep their bodies still”. An opportunity for questions was given prior to the start of the task. There were 90 Pitch Imagery Trials, 30 Pitch Perception Trials, and an average of 22 Mental Arithmetic Trials (range 14 - 40). The Pitch Perception trials were randomly interleaved with the Imagery trials after an initial 10 Imagery trials were presented. The accuracy of response for the Perception trials did not impact on the progression of the participant through the task. Mental Arithmetic trials were presented as participants moved between stages; one trial if moving up a stage or

level, and two trials if moving down a level. The average time taken to complete the task was 53 mins.

The task also included a fast exit in which participants who failed to successfully progress through Level 1 of the Imagery Trials on more than 3 attempts (that is, got more than 18 incorrect responses for Level 1 Imagery Trials) were excused from further trials. These participants were deemed to have found the task too difficult or failed to understand how to complete it. At each point of failing Level 1, the participants were given the opportunity to ask questions and the requirements of the task were reiterated verbally.

Upon completion, participants were visually presented with their percent correct scores for each Imagery level, as well as overall percent correct for the Perception and Mental Arithmetic conditions. They were then asked verbally to rate from 1 – 5 overall how vividly or clearly they formed the musical images during the task (1 – not at all vivid; 5 – very vivid) (Herholz et al., 2008; Leaver et al., 2009). They were also asked: "What strategies did you use to complete the musical imagery task?" Responses were written down and later categorised into one of several groups during analysis. Participants then completed a musical experience questionnaire which included questions of past and current, formal and informal musical participation, as well as the BAIS.

3.3 Results

Four participants failed to progress past Level 1 and were excluded from the final analyses. All four excluded participants were non-musicians with an average Musical Experience Index of .04, that is had actively participated in playing music for 4% of their lives (range 0 – .12).

3.3.1 Overall Accuracy and Reaction Times

A 2 x 3 ANOVA of Accuracy (Group: Musician, Non-Musician) x Condition (Imagery, Perception, Mental Arithmetic) revealed a significant main effect for Condition, ($F_{(2,102)} = 4.46$, $p = .01$, $\eta^2 = .07$) but no significant main effect for Group ($F_{(1,102)} = 0.04$, $p = .83$). The Group x Condition interaction was also significant ($F_{(2,102)} = 6.58$, $p = .002$, $\eta^2 = .10$), due to the fact that musicians were more accurate than non-musicians on the Imagery (musicians: M = .820, SD = 0.09; non-musicians: M = .763, SD = 0.05; $t(34) = 2.07$, $p = .046$, $d = 0.75$) and Perception conditions (musicians: M = .906, SD= 0.11; non-musicians: M = .833, SD = 0.14 ; $t(34) = 1.714$, $p = .096$, $d = 0.61$), while non-musicians were more accurate than musicians in the Mental Arithmetic condition (musicians: M = .795, SD = 0.16; non-musicians: M = .910, SD = 0.06; $t(34) = 2.351$, $p = .025$, $d = 0.86$).

A 2 x 3 ANOVA of Mean Hit Reaction Times (Group (Musician / Non-Musician) x Condition (Imagery; Perception; Mental Arithmetic)) revealed a significant main effect of Group ($F_{(1,102)} = 6.167$, $p = .02$, $\eta^2 = .05$), with musicians showing slower overall reaction times than non-musicians. There was also a significant main effect of Condition ($F_{(2,102)} = 5.034$, $p = .008$, $\eta^2 = .08$). Post hoc paired t-tests showed that reaction times (ms) were not significantly different for the Imagery (M = 1027.9, SD = 215.6) and Perception conditions (M = 992.3, SD = 250.6), but differed significantly between Imagery and Mental Arithmetic (M = 845.1, SD = 317.6): ($t(35) = 4.92$, $p < .001$, $d = 0.82$), as well as Perception and Mental Arithmetic: ($t(35) = 3.38$, $p = .002$, $d = 0.56$). All post hoc tests used the Bonferroni correction procedure with a critical alpha of .05/3. There was no significant interaction Group x Condition interaction ($F_{(2,102)} = 0.589$, $p = .56$, $\eta^2 = .01$). Figure 3-2 shows a summary of these results.

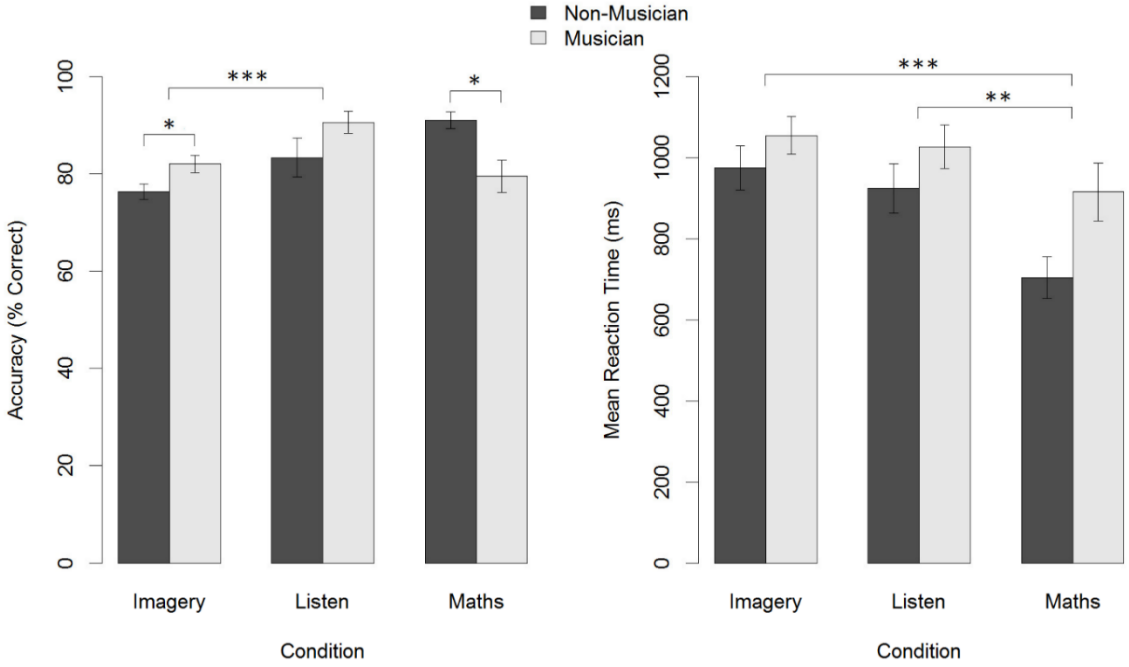

Figure 3-2: Accuracy and reaction time on the PIAT. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

3.3.2 Bucknell Auditory Imagery Scale

BAIS-V and BAIS-C were correlated across the three conditions for overall accuracy (percent correct) and mean hit reaction times. BAIS-V was significantly correlated with Imagery accuracy ($r = .49$, $p = .002$). BAIS-C was significantly correlated with overall accuracy for the Imagery ($r = .59$, $p < .001$) and the Perception condition ($r = .33$, $p = .049$) as well as significantly negatively correlated with the mean hit reaction times for both the Imagery condition ($r = -0.51$, $p = .001$) and the Perception condition ($r = -0.39$, $p = .019$).

The BAIS measures were also correlated significantly with a number of other variables. The participant's debriefing vividness rating was significantly correlated with BAIS-V ($r = .51$, $p = .001$), though not with BAIS-C ($r = .25$, $p = .15$).

The MEI showed a significant positive correlation with BAIS-C (r = .41, p = .014), but not with BAIS-V (r = .24, p = .17). However an independent t-test revealed musicians scored significantly higher than non-musicians only on BAIS-V; (musicians: M = 5.40, SD = 0.77; non-musicians: M = 4.71, SD = 1.13; $t(34) = 2.16$, p = .038, d = 0.76), not BAIS-C (musicians: M = 5.53, SD = 0.91; non-musicians: M = 5.03, SD = 0.88; $t(34) = 1.57$, p = .125, d = 0.55). Table 3-3 shows the correlation matrix for all variables of interest.

3.3.3 Strategy Use

An open-ended question asked participants to describe the strategies they had used for performing the imagery task. Responses fell into two broad categories, *Musical Imagery*, or *Alternative Strategy*. Musical imagery strategy users (N = 21) reported hearing the sound in their head or singing the notes in their head thereby following the arrows and hearing the sounds in their minds throughout the imagery (continuation) component of the task. Alternative Strategy users (N=14) reported a variety of alternative cognitive strategies such as counting arrows, using intuition or visual imagery. These participants were keeping track of the movement of the arrows but using musical imagery only at the end of the trial, to make a judgement regarding whether the test tone was correct. As the goal of the PIAT is to induce musical imagery throughout the imagery component of the trial, this later group was classified as using an alternative strategy; with only minimal musical imagery induced.

Table 3-3

Correlation Matrix of key variables.

	% Correct Imagery	% Correct Perception	% Correct Maths	Ave RT Imagery	Ave RT Perception	Ave RT Maths	BAIS Vividness	BAIS Control	Debrief Vividness	Musician Y/N	MEI	Strategy Used	Max Level
% Correct Imagery	1.000												
% Correct Perception	0.488 **	1.000											
% Correct Maths	0.075	0.134	1.000										
Ave RT Imagery	-0.479 *	-0.219	-0.253	1.000									
Ave RT Perception	-0.409	-0.478 *	-0.252	0.672 ***	1.000								
Ave RT Maths	-0.337 **	-0.171	-0.674 ***	0.735 ***	0.639 ***	1.000							
BAIS Vividness	0.555 **	0.271	-0.240	-0.203	-0.064	0.046	1.000						
BAIS Control	0.674 ***	0.353 *	-0.088	-0.51 **	-0.380 *	-0.309	0.713 ***	1.000					
Debrief Vividness	0.497 ***	0.358 *	-0.133	-0.234	-0.197	-0.073	0.546 **	0.281	1.000				
Musician [Y/N]	0.315 *	0.268	-0.40 *	0.165	0.175	0.335	0.357 *	0.275	0.331 *	1.000			
MEI	0.483 **	0.370 *	-0.171	-0.152	-0.194	-0.025	0.237	0.408 *	0.307	0.710 ***	1.000		
Strategy Used	0.614 ***	0.604 ***	0.108	-0.313	-0.415 *	-0.255	0.270	0.496 **	0.517 **	0.270	0.549 ***	1.000	
Max Level	0.791 ***	0.519 ***	-0.030	-0.382 *	-0.300	-0.161	0.484 **	0.475 **	0.576 ***	0.323 *	0.458 **	0.650 ***	1.000

Significance is denoted as * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

In addition, all participants who reached above Level 4 were asked if they possessed absolute pitch (AP), of which two self-reported they did. One reported that although they possessed AP they were not labelling the notes, just imagining the sound in their mind; hence they were categorised as using a musical imagery strategy. The other used an unusual alternative visual-motor strategy. This later individual was the only participant to reach above level 4 on the PIAT without the use of a musical imagery strategy. This participant was excluded from further analysis of strategy use and is considered in more detail in the discussion section. As it could be argued that AP possessors are likely to have clear long-term mental categories which are highly likely to influence performance on the task, the following analyses were later re-run excluding both AP possessors, but it had no effect on any of the current results; as such the AP possessor who reported using musical imagery remains included in the results below.

Musical imagery strategy users were significantly more accurate on both Imagery ($t(33) = 4.46$, $p < .001$, $d = 1.54$) and Perception ($t(33) = 4.35$, $p < .001$, $d = 1.50$) trials. Significantly faster mean hit reaction times were found for musical imagery strategy users in the Perception condition ($t(33) = 2.62$, $p = .013$, $d = 0.90$), though not the Imagery or Mental Arithmetic condition.

While musicians did not differ significantly from non-musicians in strategy use ($t(33) = 1.61$, $p = .117$, $d = 0.56$), there was a significant correlation between the MEI and strategy used ($r = .55$, $p < .001$), such that those with greater musical experience over their lifetime were more likely to report using a musical imagery strategy.

Finally, an independent t-test revealed that musical imagery strategy users reported significantly higher BAIS-C (musical imagery: M = 5.75, SD = 0.86; alternative strategy: M = 4.82, SD = 0.76; $t(33) = 3.28$, $p = .002$, $d = 1.13$), though there was no significant difference on BAIS-V (musical imagery: M = 5.38, SD = 1.08; alternative

strategy: M = 4.86, SD = 0.66; $t(33) = -1.61$, $p = .12$, $d = 0.56$). They also had significantly higher debrief vividness scores (musical imagery: M = 3.93, SD = 0.84; alternative strategy: M = 2.96, SD = 0.75; $t(33) = 3.47$, $p = .001$, $d = 1.20$).

3.3.4 Additional Imagery Performance Measures

The maximum level reached in the PIAT corresponds to the number of tones imagined per trial prior to the test probe. Both BAIS-V ($r = .46$, $p = .005$) and BAIS-C ($r = .44$, $p = .007$) were significantly correlated to maximum level attained. Musicians attained a significantly higher maximum level of performance than non-musicians (musicians: M = 4.04, SD = 1.00; non-musicians: M = 3.33, SD = 0.89; $t(34) = 2.08$, $p = .045$, $d = 0.73$). Maximum level was also correlated significantly with the MEI ($r = .40$, $p = .015$). Musical imagery strategy users reached a significantly higher level on the PIAT than alternative strategy users (musical imagery: M = 4.79, SD = 0.98; alternative strategy: M = 3.45, SD = 0.34; $t(33) = 5.781$, $p < .001$, $d = 1.69$). Maximum level reached is a more useful measure of imagery accuracy than Imagery percent correct, which does not account for the variability of difficulty in the levels of the task.

To capture how rapidly participants progressed up through the levels (i.e. whether repeated mistakes caused them to drop a level, or whether they progressed up swiftly through to Level 5 and remained there), 'Rate of Progression' was calculated as the slope of the line of best fit of the level number over the 90 Imagery trials, (setting the intercept at trial 1 as Level 1). An independent t-test between the musicians and non-musicians revealed no significant difference in this Rate of Progression (musicians: M = 0.04, SD = 0.02; non-musicians: M = 0.03, SD = 0.02; $t(34) = 1.92$, $p = .06$). However, there was a significant correlation with MEI (r=.44, $p = .008$). Musical imagery strategy users progressed significantly faster than alternative strategy users ($t(33) = 4.56$, $p < .001$). Rate

of Progression also correlated significantly with both BAIS-V (r = .45, p = .006) and BAIS-C (r = .43, p = .009).

The rate of change in reaction time as a participant moved through stages and levels of a condition provides an index of how quickly participants improved and is also an indication of the relative difficulty of the three conditions. ANOVA confirmed significant overall differences in difficulty ($F_{(2,105)}$ = 6.24, p = .003, η^2 = .11). Post hoc paired t-tests showed that Imagery (M = -8.11, SD = 6.10) was significantly more difficult than Perception (M = -31.27, SD = 28.12): (t(35) = 5.62, p < .001, d = 0.94), but there was no significant difference between the Mental Arithmetic condition (M = -20.32, SD = 38.67) and the other two conditions (See Figure 3-3).

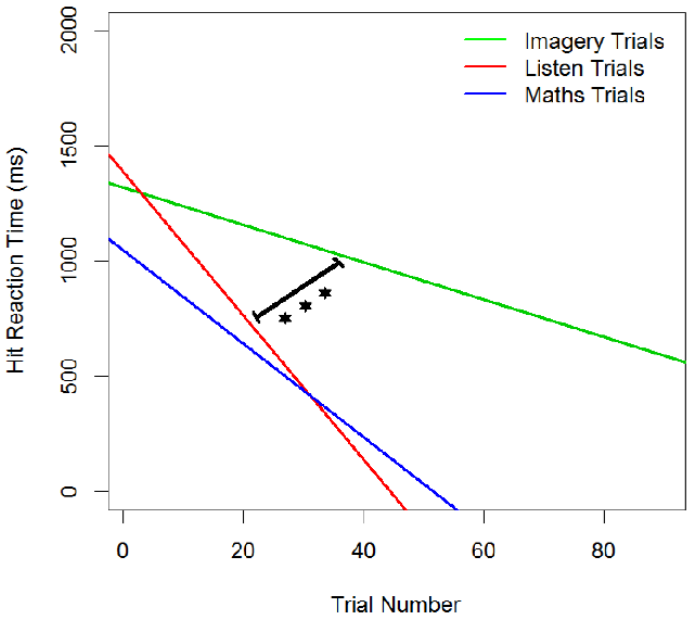

Figure 3-3: Rate of change of Reaction Time over trials

3.3.5 Multiple Regression Analysis

Regression analyses were conducted to evaluate which variables best predicted accurate performance on the PIAT, taking maximum level reached as the criterion variable. Eleven predictor variables (overall accuracies for Perception and Maths

conditions; reaction times for Imagery, Perception, Maths Conditions; Musician [Y = 1, N = 0]; MEI; Strategy Use [Musical Imagery = 1; Alternative Strategy = 0]; BAIS-V; BAIS-C; Debrief Vividness) were included in an initial model and stepwise regression reduced the model to the minimal adequate hierarchical linear model, with only significant predictors. This resulted in a final model containing strategy use and BAIS-V, which significantly predicted the maximum level reached ($F_{(2, 32)} = 17.69$, $p < .001$), and accounted for 53% of the variance in the maximum level score ($R^2 = .525$; $R^2_{adj} = .495$). The coefficients from this model are outlined in Table 3-4, under Model 2.

A linear regression calculated the variance attributable to strategy use alone, given the high correlation between strategy use and maximum level (see Table 3-3). The result was significant ($F_{(1, 33)} = 24.11$, $p < .001$), with strategy use alone accounting for 42% of the variance in maximum level ($R^2 = .422$; $R^2_{adj} = .405$). An ANOVA revealed that Model 2 (Strategy Use + BAIS-V) was significantly better than Model 1 (Strategy Use alone) ($F_{(2, 32)} = 6.935$, $p = .013$), as seen Table 3-4.

Additional linear regressions assessed the effect of adding in musical training. Model 3 and 4 on Table 3-4 show the addition of musician category group and MEI respectively. Neither model was a significant improvement, with R^2 increasing from Model 2 by only .003 and .008 respectively. This suggests that maximum level reached is either not strongly predicted from musical experience (only from Strategy Use and BAIS-V), or that the influence of strategy use or BAIS-V are mediating the relationship between performance of the PIAT and musical training.

Table 3-4

Summary of multiple regression analysis for variables predicting Maximum Level Reached (N = 35)

	Model 1			Model 2			Model 3			Model 4		
Variable	**B**	**SE B**	**β**	**B**	**SE B**	**β**	**B**	**SE B**	**β**	**B**	**SE B**	**β**
Strategy Use	1.341	0.273	0.642***	1.155	0.261	0.553***	1.131	0.27	.552***	1.044	0.308	.50**
BAIS-V				0.356	0.135	.325*	0.336	0.144	.307*	0.346	0.137	.315*
Musician [Y=1 / N=0]							0.133	0.287	0.061			
MEI										0.474	0.676	0.1
R^2_{Adj}; R^2; ΔR^2	.405; .422; .422			.495 .525; .103			.483; .528; .003			.487; .533; .008		
F for change in R^2	F[1,33] = 24.11***			F[2,32] = 6.935*			F[3,31] = 0.215			F[3,31] = 0.492		

Significance is denoted as * = $p < .05$, ** = $p < .01$, *** = $p < .001$

To test these possibilities, mediation analysis was run using MEI as the predictor, strategy use as the mediating variable and maximum level reached as the outcome variable. Logistic regression confirmed MEI significantly predicts strategy use ($z = 2.817$, $p = .005$), and linear regression confirmed strategy use significantly predicts maximum level ($\beta = 0.64$, $p < .001$). The direct effect of MEI predicting maximum level went from significant ($\beta = 0.45$, $p = .006$) to non-significant when controlling for strategy use ($\beta = 0.14$, $p = .37$), suggesting the mediation was substantial. This result suggests that musical training, though related to the performance on the PIAT (as measured by the maximum level reached), is only predictive of performance due to the impact it has on strategy use. MEI did not significantly predict BAIS-V ($\beta = 0.24$, $p = .171$), and so it can be ruled out as a mediating factor in the relationship between musical training and performance.

Finally, the role of BAIS-C was investigated to see how it fits into this model of prediction. Logistic regression showed that BAIS-C significantly predicted strategy use ($z = 2.635$, $p = .008$). The direct effect of BAIS-C predicting maximum level went from significant ($\beta = 0.44$, $p = .007$) to non-significant when controlling for strategy use ($\beta = 0.20$, $p = .19$), again suggesting that strategy use was substantially mediating the relationship between BAIS-C and maximum level reached. Linear regression also confirmed BAIS-C significantly predicted BAIS-V ($\beta = 0.71$, $p < .001$), but when controlling for BAIS-V, both variables were no longer significant in predicting maximum level reached. Figure 3-4 describes the final model for maximum level reached showing strategy use and BAIS-V as the main predictors, and MEI and BAIS-C separately predicting strategy use (though not when combined), and BAIS-C also predicting BAIS-V.

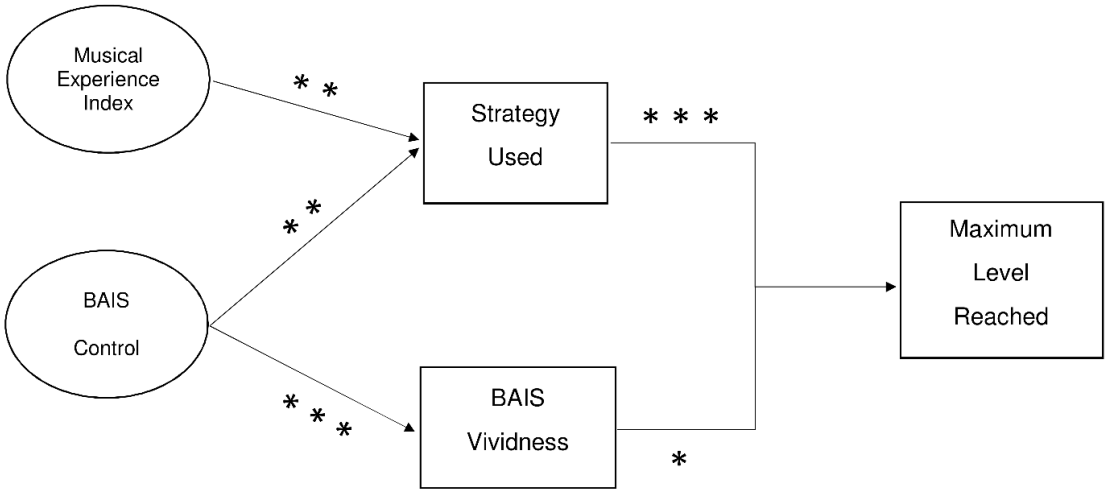

Figure 3-4: Regression model of maximum level of performance on the PIAT

In summary, musicians were significantly more accurate than non-musicians for Imagery but not Perception trials, and there was no significant group difference in

reaction times. BAIS-C scores were positively correlated with performance on both Imagery and Perception trials, as well as other measures such as strategy used and MEI. In contrast BAIS-V was positively correlated with Imagery accuracy and debrief vividness. Regression analysis showed that the factors that contributed most to better performance on the PIAT were strategy use and BAIS-V. Participants with more musical experience were more likely to use a musical imagery strategy, and therefore perform better at the task. BAIS-C also predicted both strategy use and BAIS-V. Figure 3-5 summarises the relationship between the main variables graphically and shows the differences in maximum level reached and both BAIS subscale scores, between the two strategy use categories. The size of the point on the graph is in proportion to the MEI, such that the larger points indicate a greater amount of life years spent participating in musical activity. It is interesting to note from Figure 3-5 that within the subset of 21 participants who used a musical imagery strategy, musical experience (MEI) did not predict the maximum level attained (r = .14, p = .536). However, there is a significant relationship between BAIS-V and maximum level attained (r = .53, p = .013), though not with BAIS-C and maximum level attained (r = .36, p = .105).

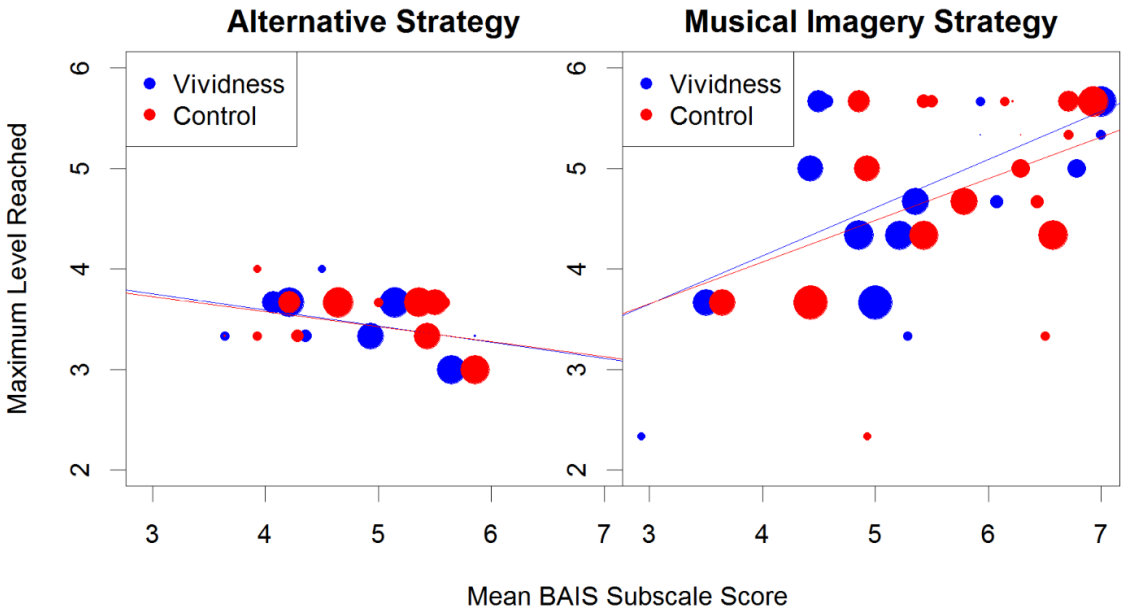

Figure 3-5: Relationship between BAIS scores and maximum level of performance for the different strategy groups. Larger circles indicate greater musical experience.

3.4 Discussion

The present results confirm the effectiveness of the PIAT for inducing musical imagery, and provide insights into the roles that musical training, auditory imagery vividness and mental control play in successful imagery performance. The PIAT was highly effective in inducing mental imagery of musical pitch in participants, in that successful performance on the task was highly dependent on the use of a musical imagery strategy, rather than an alternative strategy that used minimal musical imagery. Participants using an alternative cognitive strategy were (bar 1) unsuccessful at reaching above Level 4 on the Imagery task, with musical imagery strategy users significantly outperforming alternative strategy users on all measures of accuracy. The exception, a participant with 10 years of piano training reported, reached level 5 using a visual-motor imagery strategy. This individual reported visualising the keys on a piano and playing them with their hand. However, this individual also possessed AP and knew which pitches

and scale degrees were being played. As the only individual to adopt a visual-motor imagery strategy, this participant was considered individually and excluded from the statistical analysis of strategy use and multiple regression analyses. The participant achieved 98% accuracy for the Imagery condition and was below one standard deviation of the sample's mean for BAIS scores on both subscales. The other participant who also self-reported having AP, reported using pitch imagery. This participant was within one standard deviation of the mean for MEI, BAIS-V and BAIS-C among musical imagery strategy users, and her exclusion from the analyses did not significantly alter any of the findings (all *p* values remained within the stated significance level). Hence, this second AP participant was included in the analyses.

In comparison to previous musical imagery tasks, the present procedure has several advantages. First, it provides a number of objective and complementary behavioural measures of accuracy (percent correct per condition, maximum level attained, rate of progression through the levels) and indices of reaction time (mean hit reaction time and rate of change of reaction time per condition). These behavioural measures revealed that better performance on the PIAT was associated with the use of a musical imagery strategy. The measures also showed musicians were only significantly more accurate on the Imagery but not on Perception trials and were not significantly different in reaction times. This result is consistent with the findings of Aleman et al. (2000), who had participants mentally compare pitches of notes corresponding to lyrics taken from familiar songs. The pattern of results also suggests the PIAT is not biased towards musicians, unlike the maintenance paradigm used by Kuchenbuch et al. (2012), in which non-musicians are significantly worse than musicians on perception trials, and were at chance level for the imagery trials.

The second advantage of the PIAT is that it requires participants to actively manipulate a pitch image, rather than just maintain it. This type of manipulative pitch imagery investigation has only up until now been done with more difficult tasks such as mental reversal of melodies or simpler pitch transposition of melodies (Foster et al., 2013; Zatorre, 2012; Zatorre et al., 2010).

Third, unlike previous protocols for inducing imagery, random sequences were manipulated in the imagery component of the trial. This design confers variety and flexibility to help minimise confounds, such as a familiarity with a melody / or familiarity with a probe combination that could be learnt over a task. For example, Level 1, Stage 1 (with number of initial arrows / tones set to 3, starting note of tonic, in the key of C Major, and with only 1 imagined arrow) had 14 different possible combinations. This number increased dramatically as the participants move through the levels and stages of the task. Not only did the pitch sequence vary randomly, but the length of the initial set-up sequence varied randomly, so that participants unaware when the imagery component of the trial would begin.

Fourth, the staircase design allowed participants to progress through levels at their own rate, while at the same time accommodating individuals with a wide range of musical experience. While 4 non-musicians did fail to progress past Level 1, 3 non-musicians made up the group of 18 who progressed past Level 4 on the PIAT. One of these non-musicians also had a MEI of 0 indicating no musical participation at all. As expected, musicians performed better on the task, with 96% of them getting to Level 3 or above. Nonetheless, 63% of the non-musicians were also able to attain a maximum level of 3 or above, confirming that the PIAT can be used to induce pitch imagery in both musicians and non-musicians.

Finally, the basic design of the PIAT allows for very comparable control conditions which do not require the use of explicit musical imagery. In the perception condition, the participants were only made aware when the probe screen appeared that the trial did not require imagery, and therefore was a Perception trial. This ensured participants were actively listening throughout the trial, in anticipation that imagery may be required at any point. It also provided an identical probe presentation for direct comparison to the Imagery trials. Importantly, the mean reaction times for Imagery and Perception trials were not significantly different (even though accuracy was higher in Perception trials), due to the use of identical probes in the two conditions.

Mental arithmetic provided a second control condition that required no musical imagery but did require increasing mental capacity as the levels increased and the calculations became longer. The difficulty of the mental arithmetic trials were comparable to the imagery trials, given the similar accuracy measure scores, and decrease in reaction time over the task. These results suggest participants learned the two tasks at a similar rate. Interestingly, musicians performed more poorly than non-musicians on mental arithmetic trials, but the reasons for these differences are unclear. Musical imagery strategy users were generally faster than alternative strategy users (as seen in the negative correlations between reaction times (RT) and strategy use in Table 3-3), and musicians using an alternative strategy (N = 7), appear to be responsible for the overall slower reaction times by musicians.

A second aim was to investigate the relative importance of musical training, imagery vividness and mental control in musical imagery. Regression analysis showed that musical experience did not contribute significantly to a linear model of prediction for maximum level reached in the PIAT. Further mediation analysis showed that the relationship between MEI and maximum level reached in the PIAT was substantially

mediated by strategy used; such that more musical activity over the lifetime increased the likelihood of a musical imagery strategy being used in the PIAT. It was the use of such a strategy that led to better performance, rather than simply musical experience. These results, though surprising, may be due to the nature of the imagery task. Studies have shown in the visual domain that object imagery (maintenance) and spatial imagery (manipulation) had differing behavioural and psychometrical properties with visual artists excelling at object imagery and scientists excelling at spatial imagery (Kozhevnikov et al., 2005). A similar discrepancy in the auditory domain may be seen in expertise among musicians and non-musicians, with musicians performing better at maintenance than manipulation of musical images.

Both BAIS subscales, though highly correlated to each other, were significantly correlated with different variables; confirming they index at least partially different aspects of the auditory imagery experience. BAIS-V correlated with the vividness rating participants gave after completing the PIAT (debrief vividness), suggesting that this more abstract auditory scale is associated with the subjective experience of musical imagery vividness during the PIAT. However it was the BAIS-C that correlated most significantly with Imagery and particularly Perception performance, suggesting that being able to manipulate sound images at will may be assisting with the anticipation of the perception of them. Pfordresher and Halpern (2013) also showed a significant relationship between a perception task involving judgement about the relative height of two tones, and BAIS-C. However, this is the first study to show a significant correlation between imagery performance and BAIS-C, presumably because the PIAT involves manipulation or change of the pitch image, which requires greater mental control than maintenance paradigms. The three main variables of interest (MEI, BAIS-V and BAIS-C) all correlated with the various accuracy measures of the PIAT. However only the BAIS-C was significantly

correlated with imagery reaction times, indicating this measure is most strongly associated with overall imagery performance in this task.

A more complete picture of the relationships between these variables emerged through regression analysis, with strategy use proving to be the biggest single predictor of maximum level reached on the PIAT. Figure 3-5 shows graphically that within the musical imagery strategy users, MEI is not as big a predictor of maximum level reached; with some participants with very little musical experience able to attain a higher level than others with considerable musical experience, but with lower BAIS-V and BAIS-C scores. The combination of strategy use and BAIS-V accounted for 53% of the variance within the maximum level reached. Although MEI and BAIS-C individually predicted strategy use, and BAIS-C predicted BAIS-V, their addition to the regression model was not a significant improvement. Hence BAIS-V is more important to predicting performance on the PIAT than MEI or BAIS-C.

Therefore, while BAIS scores have been shown to be more important than musical experience in performance on this musical imagery task, it is clear the most important factor is the use of a musical imagery strategy. Interestingly, participants who reported using a musical imagery strategy were more accurate and had faster reaction times for Perception trials also, indicating that even when no manipulation of an auditory image was required, performance was facilitated by a musical imagery strategy.

It could be argued that the use of up and down arrows in the PIAT reflects a spatial conception of pitch that may encourage the use of a spatial imagery strategy to complete the task. However, arrows were presented merely to indicate which pitch to imagine next, and participants were explicitly instructed to imagine the sounds of the pitches. Indeed, any learned association with pitch height could have been used to guide imagery. We asked participants to describe the strategies they used in completing the musical

imagery trials, and the most common and successful of strategy was musical imagery rather than visual imagery. More generally, the PIAT can be readily adapted to other culture-specific schemata. For example, the major scale reflects a western conceptualization of pitch, but the PIAT can easily be modified to alternate musical scales (e.g., pentatonic, slendro, whole tone).

3.5 Conclusions

In this investigation, the PIAT was introduced as a powerful new protocol for assessing musical imagery. We confirmed that the PIAT reliably induces pitch imagery in individuals with a range of musical experience, particularly above Level 4. It entails the active manipulation of an auditory image that most non-expert musicians can readily perform. Our results showed competent performance on the PIAT requires active musical imagery and is very difficult to achieve using alternative cognitive strategies.

The PIAT provides a platform in which to address questions of individual differences in musical expertise in imagery performance, as well as the role of auditory imagery vividness and mental control. More musical training, increased self-reported BAIS-V and BAIS-C were associated with better performance on the PIAT. Both BAIS subscales were important, as success in the task required more than the ability to just hear an image in the mind, but involved the ability to successfully manipulate or change that musical image. Our results also support our second hypothesis that both auditory imagery vividness and the ability to control auditory images are more important than musical training in contributing to success in this type of imagery.

The task is readily adaptable to neuroimaging studies of the neural correlates of pitch imagery. The basic protocol also lends itself to investigations of aspects of musical imagery including loudness, tempo or rhythm. For example, a future rhythm imagery task

could involve the presentation of a simple rhythmical pattern and arrows pointing either upwards/downwards to increase/decrease divisions in the beats or left/right to either mentally reverse or maintain the simple pattern. A future study could compare imagery performance on different types of pitch scales; though it is expected that both musicians and non-musicians would perform poorly when unfamiliar scales are used, and if the task is too difficult it may encourage the use of alternative strategies.

Feedback in the PIAT was included to facilitate acquisition of task performance. It could be argued that feedback could have influenced ratings of the vividness of musical imagery, in that participants might assign lower ratings of vividness should their overall imagery performance have been perceived as poor. We acknowledge this possibility but maintain that the benefits of feedback (at least during initial learning of the task) outweigh the potential disadvantages. More explicit instructions of the types of musical imagery strategies that should be used, as well as the alternatives strategies that should be consciously avoided, may also lead to a higher percentage of participants adopting the desired auditory imagery strategy.

Looking forward, the PIAT can also be used to address other theoretical issues surrounding musical imagery. First, unlike the visual domain, where it has been demonstrated that primary visual cortex is employed during visual imagery (Kosslyn & Thompson, 2003), there is debate concerning whether primary auditory cortex is involved in musical imagery (Kraemer et al., 2005; Zatorre & Halpern, 2005). Second, it is unclear how mechanisms underlying musical imagery can be integrated into current models of auditory memory (Baddeley, 2012; Cowan, 2008). Baddeley and Logie (1992) suggests auditory imagery may reside in the “phonological loop” which includes both an auditory memory store and an articulatory rehearsal process, rather than the “central executive”. Musical imagery also has practical implications. There is considerable interest in the use

of imagery and mental practice in music education, and in successful group performance (Keller, 2012; Pascual-Leone, 2003). Finally, a clearer understanding of the links between musical imagery and perception may prove beneficial for patients with hearing loss, or for post-lingual recipients of cochlear implants; who may have functional musical imagery capabilities but have reduced capacity to perceive music. For example, incorporating the PIAT in music-based training for these patients may be beneficial, particularly for individuals with higher BAIS-C scores, given its association with Perception performance.

3.6 References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Music training and mental imagery ability. Neuropsychologia, 38(12), 1664-1668. doi:10.1016/S0028-3932(00)00079-8

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. Annual review of psychology, 63, 1-29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Logie, R. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), Auditory imagery (pp. 179-197). Hillsdale, N.J.: L. Erlbaum Associates.

Bailes, F., Bishop, L., Stevens, C. J., & Dean, R. T. (2012). Mental imagery for musicial changes in loudness. Frontiers in Psychology, 3, 525. doi:10.3389/fpsyg.2012.00525

Cebrian, A. N., & Janata, P. (2010). Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. Brain Research, 1342, 39-54. doi:10.1016/j.brainres.2010.04.026

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), Progress in Brain Research (Vol. 169, pp. 323-338): Elsevier.

Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. NeuroImage, 75, 27-35. doi:10.1016/j.neuroimage.2013.02.044

Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), Auditory Imagery (pp. 1-27). Hillsdale, N.J.: L. Erlbaum Associates.

Halpern, A. R. (2015). Differences in auditory imagery self-reported predict neural and behavioral outcomes. Psychomusicology: Music, Mind, and Brain, 25(1), 37-47. doi:10.1037/pmu0000081

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia, 42(9), 1281-1292. doi:10.1016/j.neuropsychologia.2003.12.017

Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. Psychology of Music, 41(6), 779-793. doi:10.1177/0305735612452186

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. Journal of Cognitive Neuroscience, 24(6), 1382-1397. doi:10.1162/jocn_a_00216

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. European Journal of Neuroscience, 28(11), 2352-2360. doi:10.1111/j.1460-9568.2008.06515.x

Hubbard, T. L. (2010). Auditory imagery: empirical findings. Psychological Bulletin, 136(2), 302-329.

Hubbard, T. L. (2013). Auditory aspects of auditory imagery. In S. Lacey & R. Lawson (Eds.), Multisensory Imagery (pp. 51 - 76). S.I.: Springer New York.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. Perception & Psychophysics, 68(5), 829-844. doi:10.3758/BF03193705

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. Annals of the New York Academy of Sciences, 1252(1), 206-213. doi:10.1111/j.1749-6632.2011.06439.x

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? Psychological Bulletin, 129(5), 723-746.

Kozhevnikov, M., Kosslyn, S. M., & Shephard, J. (2005). Spatial versus object visualizers: a new characterization of visual cognitive style. Memory & Cognition, 33(4), 710-726. doi:10.3758/BF03195337

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. Nature, 434(7030), 158.

Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., & Pantev, C. (2012). Electromagnetic correlates of musical expertise in processing of tone patterns. PLoS ONE, 7(1). doi:10.1371/journal.pone.0030171

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *The Journal of Neuroscience, 29*(8), 2477-2485. doi:10.1523/jneurosci.4921-08.2009

Neisser, U. (1972). Changing conceptions of imagery. In P. W. Sheehan (Ed.), *The functions and nature of imagery* (pp. 233-251). New York: Academic Press.

Pascual-Leone, A. (2003). The brain that plays music and is changed by it. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 396-412). Oxford: Oxford University Press.

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic bulletin & review, 20*(4), 747-753. doi:10.3758/s13423-013-0401-8

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Music perception and imagery in EEG: alpha band effects of task and stimulus. *International Journal of Psychophysiology, 82*(3), 254-259. doi:10.1016/j.ijpsycho.2011.09.007

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Memory & Cognition, 38*(2), 163-175. doi:10.3758/MC.38.2.163

Yoo, S.-S., Lee, C. U., & Choi, B. G. C. A. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport, 12*(14), 3045-3049.

Yumoto, M., Matsuda, M., Itoh, K., Uno, A., Karino, S., Saitoh, O., . . . Kaga, K. (2005). Auditory imagery mismatch negativity elicited in musicians. *NeuroReport, 16*(11), 1175-1178. doi:10.1097/00001756-200508010-00008

Zatorre, R. J. (2012). Beyond auditory cortex: working with musical thoughts. *Annals of the New York Academy of Sciences, 1252*(1), 222-228. doi:10.1111/j.1749-6632.2011.06437.x

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron, 47*(1), 9-12. doi:10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22*(4), 775-789. doi:10.1162/jocn.2009.21239

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*(1), 29-46. doi:10.1162/jocn.1996.8.1.29

Chapter 4

Applying Modern Psychometric Techniques to the Pitch Imagery Arrow Task

Rebecca W. **Gelding**[a,c], Peter M.C. **Harrison**[d], Blake W. **Johnson**[a,c],

William F. **Thompson**[b,c], Daniel **Müllensiefen**[e]

[a] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[b] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

[c] ARC Centre of Excellence in Cognition and its Disorders

[d] School of Electronic Engineering and Computer Science, QUEEN MARY, UNIVERISTY OF LONDON, London

[e] Department of Psychology, GOLDSMITHS, UNIVERSITY OF LONDON, London

Author Contributions

Study 1: R.W.G., W.F.T. & B.W.J. conceived and designed the experiment. R.W.G. conducted the experiment. R.W.G., P.M.C.H & D.M. analysed the data. Study 2: R.W.G., P.M.C.H & D.M. conceived and designed the experiment. R.W.G. created the stimuli. P.M.C.H. implemented into online testing software. P.M.C.H & D.M. conducted the experiment. R.W.G., P.M.C.H & D.M. analysed the data. Both studies: R.W.G., P.M.C.H & D.M. discussed the results and wrote the manuscript. All authors edited the manuscript.

Abstract

The Pitch Imagery Arrow Task (PIAT) was designed to induce and evaluate pitch imagery in participants with a range of musical backgrounds (Gelding, Thompson, & Johnson, 2015). Given a tonal context and an initial pitch sequence, arrows are displayed to elicit a scale-step sequence of imagined pitches, and participants indicate whether the final imagined tone matches an audible probe. Competent task performance requires active musical imagery and is very difficult to achieve using alternative cognitive strategies. However, the original version of the task (Gelding et al., 2015) itself is quite long (N = 90 trials). Therefore, the present two-part study aimed at enhancing its validity and reliability through the use of modern psychometric techniques, including Item Response Theory (IRT) and Computerised Adaptive Testing (CAT) (P. M. C. Harrison, Collins, & Müllensiefen, 2017). First, in an exploratory study, the original PIAT was completed by 115 participants. The data were modelled using general mixed effects models to determine main predictors of item difficulty. A new item bank was then created that systematically varied these different aspects of music structure to manipulate the perceptual difficulty of items. Second, a calibration study made use of the new item bank (N = 3000 items) to test a second participant sample (N = 243), where each participant received 30 randomly selected items from the set of 3000. Using explanatory item response modelling (generalised mixed effects models) four predictors were identified as contributing significantly to item difficulty. Ability on the PIAT was found to require the ability to maintain and manipulate tones in mental imagery, as well as to resist perceptual biases that can lead to incorrect responses.

4.1 Introduction

Mental imagery is the representation in the mind of a sensory experience in the absence of sensory input (Hubbard, 2018). Though ancient philosophers such as Aristotle believed that imagination was central to thought itself (MacKisack et al., 2016), it wasn't until the 1970's that modern research began to explore the phenomenon of visual imagery (Kosslyn, 1973; Shepard & Metzler, 1971). Visual images can be subjected to a number of operations such as inspection, zooming, rotation and transformation (Thagard, 2005). However only in the 1990's was the first volume was written on the study of imagery in the auditory modality which aimed to explore analogous operations the auditory domain (Reisberg, 1992).

Musical imagery is a subset of auditory imagery and has been described as the silent mental replaying of music in one's own mind (Halpern, 2003). This ability to internally hear music has been argued to be fundamental to musical talent (Gordon, 1989b; Seashore, 1919), hence the earliest application of the study of musical imagery was limited to music education; teaching young musicians to imagine the desired sound to coordinate their movement to enable that sound to occur (Goldsworthy, 2010). Music education and ensemble playing remain an important application of musical imagery research today (Keller, 2012; Keller & Appel, 2010), however current research is also exploring application to movement disorders such as Parkinson's disease or stroke (Lee et al., 2018; Schaefer, 2017), memory disorders such as dementia (Halpern et al., 2015) as well as role of musical imagery in auditory hallucinations in clinical and non-clinical populations (Kumar et al., 2014; Linden et al., 2011; Shinosaki et al., 2003). To effectively study musical imagery in these various applications, efficient and reliable tests of musical imagery ability are required.

While numerous studies have attempted to test musical imagery abilities, particularly their neural correlates (Cebrian & Janata, 2010; Halpern, 1992; Herholz et al., 2012; Herholz et al., 2008; Leaver et al., 2009; Zatorre & Halpern, 2005; Zatorre et al., 2010; Zatorre et al., 1996), most to date have explored passive musical imagery, using paradigms requiring continuation of familiar melodies in silence (Herholz et al., 2008; Weir et al., 2015), or comparisons of pitches from lyrics of familiar songs (Aleman et al., 2000; Halpern, 1992). Active musical imagery, which requires manipulation and control over the imagined content, has received less attention (Halpern, 2012; Zatorre et al., 2010). Across both forms, several limitations in the study of musical imagery remain. These include lack of objective measures of performance (Kraemer et al., 2005); and inflexibility – tasks that are too easy for musicians (Janata & Paroo, 2006), or too hard for non-musicians (Zatorre et al., 2010). The Pitch Imagery Arrow Task (PIAT) was designed to address these limitations (Gelding et al., 2015).

The PIAT has several advantages over existing protocols for evaluating imagery. Specifically, the task (1) requires a behavioural response to objectively measure accuracy and response times of imagery performance; (2) is extremely difficult to successfully perform using alternative cognitive strategies other than pitch imagery; (3) employs novel rather than familiar sequences of pitches that cannot be anticipated in advance; (4) employs a range of difficulties implemented in a staircase design, such that it can induce imagery in participants with a wide range of musical experience. However, one of the main disadvantages is the time taken to complete the task. With 90 trials, the task is time consuming and experienced as tedious by many participants.

One way to optimise tests of individual differences, making them more time-efficient and reliable, is through modern psychometric techniques such as Item Response Theory (IRT), and Computerised Adaptive Testing (CAT) (P. M. C. Harrison et al., 2017).

The main prerequisite for a PIAT version using IRT and CAT, is a psychometric model that predicts the difficulty of PIAT items. The aim of the present studies was to construct such a model. Firstly, an exploratory study using the original PIAT, tested 115 participants to determine the key variables that contribute to item difficulty. A cognitive model of the processes used to complete a PIAT trial was then developed on the basis of these exploratory results. Subsequently, a calibration study was conducted that systematically tested a large bank of pre-generated items and determined parameters of an explanatory IRT model. This final model serves to construct a future computerised adaptive version of the PIAT.

4.2 Study 1: Exploratory Phase

The aim of the first study was to identify features of musical structure and aspects of trial design that contribute to item difficulty on the original PIAT and hence to generate an initial psychometric model of task performance on the PIAT.

4.2.1 Materials and Methods

4.2.1.1 Participants

A total of 115 participants completed this study over three recruitment stages. The initial 40 participants (22 females) were recruited for the original PIAT study (Gelding et al., 2015). An additional 24 participants (15 females) completed an identical task as outlined in Gelding et al. (2015), in order to qualify for a different study. All of these participants (n = 64) completed the original version of the PIAT along with 2 control conditions – perception and mental arithmetic. The remaining 51 participants (35 females) completed the PIAT with only imagery trails included (that is, no mental arithmetic or perception control conditions). This latter group also completed a rhythm imagery task during the experimental session either before or after the PIAT.

4.2.1.2 Materials

Pitch Imagery Arrow Task (PIAT)

An individual trial on the PIAT begins with an ascending major scale to provide a tonal context. A start note (either tonic or dominant of scale) is then presented simultaneously with the visual presentation of a dot on the screen. A variable number of up / down arrows are next displayed in random order, with each arrow accompanied by a corresponding pitch that moves up / down the scale in a stepwise motion. Pitch changes always match the direction indicated by the arrows. These stimuli are followed by a continuation phase consisting of a number of silent arrows, in which participants are required to imagine the corresponding changes in pitch. Immediately after the sequence of silent arrows, a pre-probe screen appears, to allow participants time to consolidate their current pitch image and prepare to hear the probe. One second later, an audible probe pitch is sounded. Participants are then required to indicate whether the probe matches the final imagined tone. When the probe is incorrect, it is always within the same key signature, so that it is not obviously wrong, and a maximum of two steps away from correct answer. A staircase design was used in which all participants began on the easiest difficulty and progressed to increased complexity with accurate responses (2 correct answers or 90% correct on a given stage of the task). See Gelding et al. (2015) for more details of the staircase design.

Psychometric Questionnaires

As well as completing the PIAT, participants also completed two questionnaires, one to measure musical background and the other to measure auditory imagery vividness and control. Firstly, participants in the first two recruitment stages (n = 64) completed a generic musical background survey, from which the years of active musical engagement was calculated. Participants from the third recruitment stage (n = 51) completed the

Goldsmith's Musical Sophistication Index (Gold-MSI; Mullensiefen et al., 2014) to obtain a comprehensive profile of their musical skills and experiences. The musical training subscale of the Gold-MSI is of particular importance for the current study given the assumed link between the ability to imagine music and the amount of formal musical training received (Aleman et al., 2000). Participants in this study showed a good spread of musical training background with scale scores ranging from 10 to 44 (mean = 26.5, median = 27, SD = 10.46) which is similar to the distribution of musical training in the general population (median = 27 in Mullensiefen et al., 2014). In addition, a Musical Experience Index (MEI) was calculated as years of musical engagement / age for participants 1 – 63 and approximated for participants 64 – 115 as years of formal training or active practice (whichever was greater) / age. Note the maximum years that could be approximated for later group of participants was 10 years, so maximum MEI for this group is 10/18 = 0.55.

Secondly, all participants completed the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015). This 7-point Likert scale includes two subscales, for vividness (BAIS-V) and control (BAIS-C), both of which have 14 items each. Participants in this study showed a range of vividness scores from 2.85 to 7 (mean = 5.025, median = 4.929, SD = .960) and a range of control scores from 3 to 7 (mean = 5.202, median = 5.286, SD = .964), which is similar to the distribution of Halpern (2015) who found both BAIS-V and BAIS-C had mean scores of 5.1 and SD of 0.9.

4.2.1.3 Procedure

Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA) was used to control the experiment and to record responses. Acoustic stimuli were generated from the 'Piano' instrument sound by Finale 2012 software (Makemusic Inc; Eden Prairie, MN) and exported as .wav files for use in Presentation®.

Upon being seated in front of the computer with headphones, participants were given a sound check, whereby they could manually adjust the volume of the tones to a suitable level. They were then introduced to the task. Participants were informed that no movement or humming was allowed, to assist them with the task, but they should "as vividly as possible, imagine the tones and keep their bodies still". An opportunity for questions was given prior to the start of the task.

The task has a fast exit in which participants who failed to successfully progress through Level 1 of the Imagery Trials on more than 3 attempts (that is, got more than 18 incorrect responses for Level 1 Imagery Trials) were excused from further trials. Fourteen participants were triaged in this way, having completed a range between 41 and 77 trials at their point of exit. These participants were deemed to have found the task too difficult or failed to understand how to complete it. At each point of failing Level 1, the participants were given the opportunity to ask questions and the requirements of the task were reiterated verbally.

Upon completion, participants were asked verbally to rate how vividly or clearly they formed the musical images during the task (1—not at all vivid; 5—very vivid). They were also asked: "What strategies did you use to complete the musical imagery task?" Verbal responses were noted down by the experimenter. Participants then completed the BAIS and musical experience or Gold-MSI questionnaires (as per Materials section).

4.2.1.4 Ethics

All participants provided written consent and all procedures were approved by the Macquarie University Human Research Ethics Committee.

4.2.2 Results

In a first step, correct responses of each participant were summed characterising each individual's performance on the PIAT. Summed scores ranged from 41.5% to 99% correct responses with a mean of 75.2% (SD = 11.7%) and a median of 75.9% (1st quartile at 70% and 3rd quartile at 82.2%). Table 4-1 shows the correlations between PIAT sum scores and demographic as well as musical background variables. There were no significant correlations between performance on the PIAT and gender or age (p-values ≥ .62). In contrast, PIAT scores correlated substantially and significantly (all p-values < .005 after correcting for multiple comparisons using Holm's (1979) procedure) with all indicators of musical background.

Table 4-1

Correlations with Performance Accuracy

	Age	**Gender**	**MEI**	**Musical Training (Gold-MSI)**	**BAIS-V**	**BAIS-C**
N	115	115	115	51	115	115
Performance Accuracy	-.043	0.045	.534***	.498**	.324**	.386***

Significance is denoted as ** = $p < .01$, *** = $p < .001$

In particular, the correlation of performance accuracy with the aggregated number of years of active musical training or engagement (MEI) of $r = .53$ ($p < .001$) supports a linkage between musical training and musical imagery ability (Aleman et al., 2000).

In a second step, data at the level of individual trials were analysed with generalised mixed effects models using the packages *lme4* (De Boeck et al., 2011), *AICcmodavg* (Mazerolle, 2017) and *psyphy* (Knoblauch, 2014) in the statistical computing environment R (R Core Team, 2014). We used the model selection strategy

based on the corrected Akaike Information Criterion (AICc) as described in Long (2012). A summary of the best logistic regression model predicting performance accuracy dependant variable (0 or 1), having the lowest AICc, is given in Table 4-2. This model used dummy coding for coding the contrasts of categorical variables.

Table 4-2

Generalised Mixed Effects Regression Model for Performance Accuracy

Predictor	**B**	**SE**	***z***	***p***
(Intercept)	1.401	0.205	6.822	< .001***
Level	-0.357	0.054	-6.574	< .001***
ProbabilityProbe	2.926	0.297	9.848	< .001***
ProbeNoteisStartNote	-0.680	0.126	-5.407	< .001***
Stage 2	-0.070	0.110	-0.631	.528
Stage 3	0.119	0.127	0.934	.350
Stage 4	0.772	0.299	2.583	.010**

Significance is denoted as ** = $p < .01$, *** = $p < .001$

The best model included random effects for participants and items, as well as 6 fixed effects for (1) Level (i.e. the number of silent arrows), (2) the probability of the probe, (3) a binary variable indicating whether the probe note was identical to the start note of the audio-visual sequence, and 3 factors for the different Stages of the trial – that represent variability in key signature, start notes and the number of heard arrows in the set-up component of a trial (see Table 4-2). The lower asymptote (guessing level) and the upper asymptote (ceiling level) of the model were optimised given these fixed and random effects, and optimal values were identified at 0.3 (chance level) and 0.95 (ceiling). The classification accuracy of the final model was 64.9% without random effects (i.e. not

incorporating model-based ability estimates from the same participants) and 71.6% with random effects (i.e. including model-based ability estimates from the same participants).

4.2.3 Discussion

The results of the exploratory study show that there are considerable individual differences between participants on the PIAT and that task performance is significantly correlated with musical training and self-reported ability to imagine auditory material. In addition, data modelling at the individual trial level showed that meaningful factors that affect task difficulty can be identified. Results of the model evaluation demonstrate that these factors (i.e. fixed effects) explain a sizeable proportion of model accuracy (64.9%). Including personal information (i.e. random effects of participant ability) further increases model accuracy to 71.6%. The sizable contributions of individual differences on the task suggest that it is especially suitable for computerised adaptive testing.

The largest predictor of item difficulty was the number of tones that the participant had to imagine: more tones led to higher difficulty. The second largest predictor was the proportion of other items in the item bank that shared the same probe tone: less frequent probe tones led to higher difficulty. In addition, we found fewer correct responses for trials where the probe tone was identical to the first tone of the sequence, which suggests a perceptual bias when the start note is used as the probe. That is, for incorrect probes when the probe was the start note, participants were more likely to select it as correct and therefore make an error. Finally, simpler trial Stages (fixed key and start note) proved to be easier for participants.

Taken together, the results of the exploratory study suggest that it is a well-suited task for constructing an effective test of pitch imagery ability based on a rigorous item response model. Results of the exploratory study also help to construct a hypothetical

cognitive model of task performance on the PIAT, which serves as the basis for the subsequent calibration study.

4.3 Cognitive Process Model

To simplify a PIAT trial, improvements were made to probe and response component of the trial. The original PIAT involved a pre-probe screen to alert participants of the need to maintain their current image and prepare them to hear the probe, which occurred 1 second later (Gelding et al., 2015). In the updated PIAT trial, the pre-probe screen was removed, and instead the final silent arrow included the word "hold" on it and was displayed for 2 seconds instead of 1 second. A cross appears on the screen when the probe is sounded (see Figure 4-1). The participants when answered the question "Does the probe match the last imagined note?", with two buttons at the bottom of the screen ("Match" or "Non-Match") to choose from.

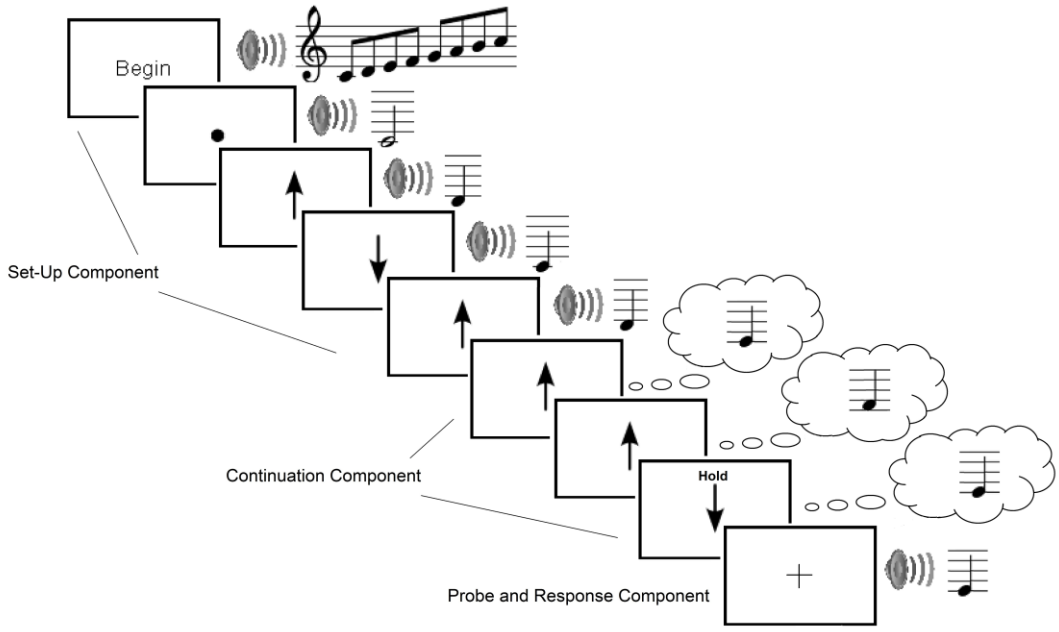

Figure 4-1: Schematic of the updated PIAT trial

Using the participant's descriptions of the strategies used to do the task, a cognitive process model was developed. The purpose of the cognitive process model was

to describe the stages of processing of a PIAT trial, to consider how different variables may be related to item difficulty, and therefore inform the future calibration modelling (P. M. C. Harrison et al., 2016). The cognitive process model included the following stages: perceptual set-up, auditory imagery generation, manipulation and maintenance, similarity comparison and decision-making (see Figure 4-2).

Perceptual set-up occurs as the participant activates the tonality template for the trials from the presentation of the initial scale and start note. Next, coordinated audio-visual processing is activated through the arrows and tones being presented together during the set-up component. Generation of the first auditory image occurs when the first silent arrow is presented. Given the uncertainty of when the first silent arrow will occur, expectation for a silent arrow increases once the initial number of heard arrows reaches 3, given that all trials had at least 3 sounded arrows in the set-up component. Subsequent processing of the silent arrows guides the manipulation of the auditory image. When the arrow with "Hold" appears, participants then maintain the last imagined note in working memory. A similarity comparison is made when the probe is heard, with a participant then making the decision whether the probe matches the last note they were imagining.

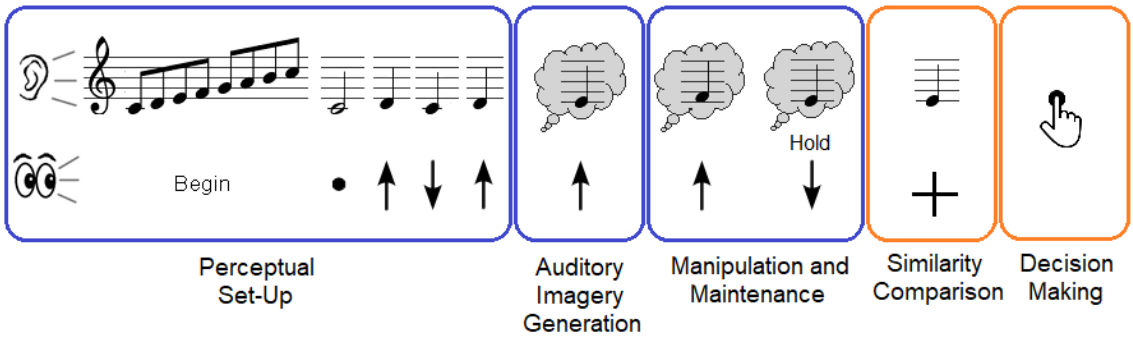

Figure 4-2: Schematic of the cognitive process model for the PIAT. Blue outlines represent processes of the model (Perceptual Set-Up, Auditory Imagery Generation, Manipulation and Maintenance) that are the same for all trials, regardless of the probe accuracy. Orange outlines represent the processes of the model (Similarity Comparison, Decision Making) that vary depending if the probe is correct or incorrect.

Item features that impair the imagery stages of the PIAT cognitive process model should increase item difficulty. For example, if the correct auditory image is not originally generated, then subsequent manipulations would lead to an incorrect response. Hence if participants fail to complete Level 1 items correctly, this suggests a lack of ability in generating a correct auditory image. Errors can also occur during manipulation: if participants are not paying full attention to the silent arrows (and lose one or more steps); if manipulations are performed incorrectly with more than a single step taken with each arrow; or if their imagery strength diminishes over the trial, leading to an impoverished or incorrect image being maintained prior to the probe. These types of errors are more likely at higher levels. In such cases participants may use the information still available to them to complete the task, some of which may cause biases in responses. For example, memory for important notes from the heard sequence (e.g. the tonic or fifth of the scale presented or indeed any note contained in sequence) may bias participants to respond as "correct" if imagery for the last note is not strong enough to compare to the probe, and the probe matches an important note from the sequence (Deutsch, 1970, 1972). This bias would increase accuracy for correct trials but result in errors for incorrect trials. Having several steps in one direction within a trial may also increase item difficulty as the correct probe would be further away from the last note heard (hence items with a larger distance between last heard note and probe may be more difficult).

Other information available to participants if they lose their imagery may be the approximate direction of the probe relative to the last note heard, which could be tracked through counting arrows. If the direction of the probe relative to the last note heard is consistent with the direction of the arrow count (i.e. if the probe is above last heard note, and arrow count is positive), then incorrect trials will be more difficult to judge, leading to increased errors. Conversely, if the direction of the probe relative to the last note heard

is inconsistent with the direction of the arrow count (i.e. if probe is above last heard note, and arrow count is negative), then incorrect trials would be much easier to detect. Incorrect trials should also be more difficult if the probe is 1 step rather than 2 steps away from the true imagined note, as the further away the probe is to the true imagined note, the more obviously wrong it will be. The final information participants may also be using in lieu of accurate imagery representations are implicit probe probability approximations, to decide on the likelihood of a given probe being correct, either based from the last note heard or the start note or the total number of arrows in the trials.

Once the probe is sounded, participants compare their imagined note with the probe and must decide whether it is correct. If the imagined probe matches the sounded probe, then a correct decision is straightforward. If it does not match, participants consider their confidence in their imagined note, and the other information at hand, to determine whether to select "incorrect" or whether they have made an error in their imagery and should instead respond as "correct". Confidence in a response should be highest when the true imagined note matches the last note heard, or when the true imagined note is the tonic or dominant of the scale. Hence this cognitive process model suggests that any explanatory model of data collected from the PIAT should consider correct trials and incorrect trials separately, and that there are many variables that can be extracted from the musical structure of a trial that could potentially predict item difficulty.

4.4 Study 2: Calibration Phase

As a result of the exploratory phase and the development of the cognitive process model, several changes were made to the PIAT and a new calibration study was conducted. The aim of the calibration study was to explore how item difficulty relates to the different features of a new set of pre-generated experimental stimuli ($N = 3000$ items). In this new set, the stimuli systematically vary on predictors identified as important in the

exploratory phase. The output of the calibration phase is an improved explanatory model that can form the basis for the computerised adaptive version of the PIAT.

4.4.1 Materials and Methods

4.4.1.1 Participants

243 participants with a mean age of 21.8 years (SD = 3.8) took part in the calibration study. 156 (65%) of the participants were female, 81 (33.3%) were male, while 3 indicated their gender as "other" and 3 preferred not to disclose their gender. 10 participants were recruited among 1st year undergraduates at Goldsmiths University of London who participated for course credit and 233 were recruited through the online panel of the consumer insights company SoundOut and received a small monetary compensation.

4.4.1.2 Materials

Trials were pre-generated as movies using open source software Openshot (www.openshot.org) and FFMPEG (www.ffmpeg.org), with piano tones from the *Alicia's Keys* piano plugin (Native Instruments GmbH) for Audacity (www.audacityteam.org). Stimuli were generated to systematically combine the variables: key signature (C, C#, D, Eb, and E Major), level (number of silent arrows in a trial from 1 – 5), number of heard arrows (3 – 5), start note (tonic or dominant), and accuracy (correct or incorrect). Hence there were 5 (keys) x 5 (levels) x 3 (heard arrows) x 2 (start notes) x 2 (accuracies) = 300 trial types. Ten variations of each trial type from a random generation of arrow combinations resulting in 3000 video stimuli being created. The only constraint was that the probe could not be the start note of the trial, and the range of notes over the trial was bounded by +/- 4 steps from the start note. This was increased from +/-3 steps from the exploratory study to increase variability and decrease the probability of a given probe.

The presentation of the stimuli was through a custom-made interface implemented using the *psychTestR* package (P. M. C. Harrison, In Preparation) and delivered through the R Shiny interface (http://shiny.rstudio.com/).

In addition, participants completed the Gold-MSI self-report questionnaire on their musical background, skills and expertise (Mullensiefen et al., 2014). Musical training of participants in this study was lower (mean = 23.37, median = 23) compared with the exploratory study. Gold-MSI musical training scores ranged from 7 to 49 with a standard deviation of 9.78.

4.4.1.3 Procedure

Participants were introduced to the task in several steps: explaining the set-up of the tonal context, the alignment of visible arrows and audible tones on the scale, and finally the silent arrows that require imagery of the corresponding tones. Participants were then presented with three practice trials where they were given feedback on their responses and were offered the opportunity to repeat the practice trials as many times as they wished. Subsequently, participants had to respond to 30 trials on the PIAT without any feedback. Items were presented quasi-randomly, with the only constraint being that participants received an equal number of items (n = 6) from each of the 5 levels. Finally, participants completed the Gold-MSI self-report questionnaire as well basic demographic questions and were asked to describe the strategy they used to complete the task by selecting one of several options.

4.4.1.4 Ethics

The experiment received ethical approval by the Ethics committee at Goldsmiths, University of London.

4.4.2 Results

The data analysis aimed to construct an explanatory item response model (De Boeck & Wilson, 2004) using a binary logistic regression models with the response score (correct / incorrect response) as dependent variable and 24 variables as potential predictors reflecting different aspects of the individual trials in terms of their musical or procedural features. The variables were derived from the cognitive process model described in Section 4.3 and a short definition of each predictor variable is given in Table 4-3.

In order to reduce the number of potential predictor variables, we performed an initial variable selection procedure employing random forest classification (Breiman, 2001) to predict the correctness (0 or 1) of the responses at the individual trial level. Random forests have the advantage that they can handle a large number of predictors and provide an index of the importance of each variable for the classification accuracy of the model. We used three different random forest variable importance measures: (1) the mean decrease in model accuracy from Breiman's (2001) original random forest implementation; (2) the mean decrease in accuracy in the conditional random forest implementation that is based on conditional inference from permutation tests of Strobl, Malley, and Tutz (2009); (3) the AUC-based variable importance measure as described by Janitza, Strobl, and Boulesteix (2013) also based on the conditional random forest implementation.

Table 4-3

Predictor variable names, definitions and range of values

Predictor Variable	Definition	Range of Values
Level	Number of imagined arrows per trial	1 – 5
Key	Key Signature (1 – 5 corresponding to C Maj, C# Maj, D Maj, Eb Maj, E Maj)	1 – 5
StartNote	Tonic (1) or Dominant (0) of scale	1 / 0
HeardArrow	Number of sounded arrows presented in Set-Up period	3 – 5
HeardRange	The number of unique tones played during the Set-Up period, including the start note.	2 – 5
ProbeStartNote Difference	Probe number relative to steps away from the start note (at 0)	-4 – +4
Probe	This is the probe note number in the scale where 1 is the tonic up to 8 which is the tonic up one octave, -4, -5, -6 and -7 are the 4th, 5th, 6th and 7th notes of the scale in the lower octave	-4 – -7; 1 – 9
ProbePrevious	Whether the probe was previously heard in the initial Set-Up period of the trial (1) or not (0)	1 / 0
LowProbe	If probe was not heard in set-up and was lower than start note (so not heard in initial scale) then (1) else (0)	1 / 0
ProbeAccuracy	Accuracy of the probe. If probe was correct = 1, if incorrect = 0	1 / 0
ProbabilityProbe	Probability of the probe, given the total number of arrows presented in the trial	0 – 0.375
ProbabilityProbe Constrained	Probability of the probe, given the total number of arrows presented in the trial and the restraint that the start note cannot be used as a correct probe	0 – 0.4
LastHeard	The last note heard in the sequence relative to start note	-3 – 3
Binomial ProbabilityProbeStart Note	Binomial probability of probe based on start note	.003 – .21
ProbabilityProbeStart Note	Probability of probe based on actual data of distance between start note and probe	.15 – .31
Binomial ProbabilityProbeLast Heard	Binomial probability of probe based on last heard note	.009 – .24
ProbabilityProbe LastHeard	Probability of probe based on actual data of distance between last heard note and probe	.0003 – .40
ProbeLastHeard AbsDiff	Absolute value of the difference between last heard note and Probe	0 – 6
ProbeNote1	Takes a value of 1 only if the probe is the tonic; is derived from Probe = 1 or 8	1/0

ProbeNote1or5	Takes a value of 1 only if the probe is the tonic or dominant; i.e. Probe = -5, 1, 5 or 8	1/0
ProbeTrueIm AbsDiff	Absolute difference between the true imagined final note and the probe presented	0,1,2
DirectionSame	Takes value of 1 only if the direction of the probe tone from the last note heard (up, down, same) is the same as the direction of the true imagine tone from the last note heard	1/0
LastHeardTrueImAbs Diff	Absolute value of the difference between last heard note and true Imagined final note	0 – 4
TrueIm1or5	Takes a value of 1 only if the true imagined final note is the tonic or dominant; is derived from true imagined note = -5, 1, 5 or 8	1/0

In accordance with the cognitive process model we allowed predictor variables to have different functions (i.e. coefficients) when modelling trials with a correct probe vs trials with an incorrect probe. To this end we created two data subsets for correct probe trials (3645 observations) and incorrect probe trials (3645 observations). Considering that the overall aim was to obtain a compact model of the data suitable as a basis for an adaptive test, we selected the ten most important predictors from each of the three random forest models for each dataset. Since the three sets of important variables showed a considerable overlap, the combined sets of the most important predictors contained twelve unique variables for each of the two datasets with correct and incorrect probe trials.

The second variable selection step made use of generalised linear mixed effects models which are able to account for individual differences by including a random effect for participants, representing participant ability. For each of the two datasets, we constructed a null model (not including any predictor variables as fixed effects, but only the random intercept effect for participants) and a full model including all predictor variables as fixed effects. The parameters for the lower asymptote (guessing parameter) and upper asymptote (inattention parameter) were optimised for each model separately. In a final step we performed an exhaustive search through all possible subsets of predictor

variables as main effects and subsequently optimised the parameters for the lower and upper asymptote again. The best model (according to the Bayesian Information Criterion, BIC) for correct probe trials contained two predictor variables (ProbabilityProbeLastHeard and Level) and had a much better fit to the data (BIC=4492.411) than the null (BIC=4509.053) and the full model (BIC=4556.578). The classification accuracy of this model was 70.8%. As per Table 4-3, ProbabilityProbeLastHeard is defined as the probability of the probe based on the actual data of distance between last heard note and the probe.

The final model for incorrect probe trials contained two predictor variables (ProbeTrueImAbsDiff and HeardRange). As per Table 4-3, ProbeTrueImAbsDiff is defined as the absolute difference between the true imagined final note and the probe presented, hence takes a value of 0 for correct trials but for incorrect trials is either 1 or 2. HeardRange is defined as the number of unique tones played during the set-up period, including the start note. This model also had a much better fit to the data (BIC=4856.6) than the corresponding null model (BIC=4903.967) and full model (BIC=4909.979). Its classification accuracy was 66.83%.

In a final step we combined the predictor variables from both models into a single model specifying an interaction effect of each predictor with the status of the probe (correct / incorrect). Coefficients for all predictors and parameters for the lower and upper asymptote were estimated on the full dataset (7290 observations). The final model had a prediction accuracy of 63.8%.

Table 4-4 provides summaries of all three models (i.e. correct probe trials, incorrect probe trials and joint model). The model summaries show that some of the predictors assume different functions for correct and incorrect probe trials. ProbeTrueImAbsDiff is only meaningfully defined for incorrect probe trials

and HeardRange has a negative coefficient for incorrect probe trials and a positive coefficient for correct probe trials. For ProbabilityProbeLastHeard only the coefficient for correct probe trails is significant. In contrast, Level has coefficients of similar magnitude for correct and incorrect probe trials, both of which are significant.

Table 4-4

Generalised Linear regression model predicting item difficulty from Correct probe trials, Incorrect probe trials and the joint model

Correct Probe Trials				
Predictor	**β**	**SE**	***z***	***p***
(Intercept)	0.839	0.155	5.411	< 0.001***
ProbabilityProbeLastHeard	1.052	0.351	2.999	0.003 **
Level	-0.116	0.027	-4.332	< 0.001***
Incorrect Probe Trials				
(Intercept)	-1.633	0.342	-4.770	< 0.001***
ProbeTrueImAbsDiff	1.028	0.135	7.607	< 0.001***
Heard_Range	-0.224	0.078	-2.886	0.004 **
Joint Model				
(Intercept)	-0.918	0.305	-3.009	0.003 **
Incorrect : ProbabilityProbeLastHeard	0.228	0.768	0.296	0.767
Correct : ProbabilityProbeLastHeard	2.778	0.600	4.627	< 0.001***
Incorrect : Level	-0.157	0.064	-2.438	0.015 *
Correct : Level	-0.176	0.046	-3.795	< 0.001***
Incorrect : ProbeTrueImAbsDiff	1.151	0.169	6.823	< 0.001***
Incorrect : HeardRange	-0.553	0.107	-5.164	< 0.001***
Correct : HeardRange	0.157	0.073	2.162	0.031 *

Note: In the Joint Model, Correct indicates ProbeAccuracy = 1, Incorrect indicates ProbeAccuracy = 0.

The selected predictors and the signs of their coefficients for correct and incorrect probe trials are consistent with the cognitive process model. Higher levels (i.e. more imagined tones) led to a lower performance which indicates that longer sequences make it more likely that participants can lose their imagery or imagine notes that are not congruent with the arrows shown. This applies to correct and incorrect probe trials alike. If participants are not able to correctly imagine the sequence of tones, they then must rely on alternative cognitive and perceptual heuristics. These heuristics include the probability of the probe given the number of arrows and the last note of the sequence heard, as well as the most salient traces in auditory memory, such as notes that were heard during the set-up sequence.

The presence of perceptual biases in responding is evidenced by the significance of the predictor ProbabilityProbeLastHeard in both the model for correct trials only and in the correct trials within the joint model, but not in incorrect trials. This variable is the probability of the probe given the last note heard, calculated from the whole dataset of 3000 items. Hence the last note heard and probe combinations that have higher probability were more likely to be selected as a "match" by participants. When the probes were correct, this means that the bias works in the participant's favour, in that they are more likely to select "match" and hence give the correct answer. However, when the probes were incorrect, and participants still select "match", this leads to an error in responding.

As predicted in the cognitive process model, for incorrect trials when the probe is 2 steps away from the true imagined note, the trials are significantly easier than when the probe is 1 step away. This is seen in the variable ProbeTrueImAbsDiff being significant both in the incorrect trial only model, and the incorrect trials in the joint model.

HeardRange is another variable which suggests the presence of perceptual biases in responding. For the incorrect trials only model, this variable significantly and

negatively predicts performance. In the joint model again for incorrect trials, the coeffieicent is significant and negative, however for correct trials in the joint model the coeffieicent is significant and positive. This suggests that when the HeardRange is larger, (i.e. the difference between the lowest and highest notes of the set-up period is greater), participants are more likely to select the probe as a "match", leading to accurate responses for correct trials, but inaccurate responses for incorrect trials.

In a last step, performance on the PIAT was compared to participants' musical background. Performance on the PIAT was measured as a sum score of correct responses as well as through the random effects coefficients extracted from the mixed effects model which represent the latent variable participant ability. Sum scores ranged from 26.7% to 100% correct responses with a mean of 59.5% (SD = 13.5%) and a median of 60% (1st quartile at 50% and 3rd quartile at 67.7%). Participant's random effects ranged from -1.78 to 3.02 with a mean of 0.057 (SD = 0.87) and a median of -0.044 (1st quartile at -0.56 and 3rd quartile at 0.56). The correlation between these two indicators of performance was r = .94. Table 4-5 shows the correlations between PIAT sum scores (performance accuracy) and random effects ability scores (participant ability) with demographic as well as musical background variables. There were no significant correlations between performance on the PIAT and gender or age (p-values $\geq$ 0.38). In contrast, PIAT scores correlated substantially and significantly (all p-values $<$ 0.001 after correcting for multiple comparisons using Holm's (1979) procedure) with self-reported perceptual abilities, emotional musical engagement and musical training. However, no significant correlations were found with self-reported active engagement, singing abilities or general sophistication.

Table 4-5

Correlations with Performance Accuracy and latent variable Participant ability

	Age	Gender	Active Engagement	Emotions	Musical Training	Perceptual Abilities	Singing Abilities	General Sophistication
Performance Accuracy	-0.037	-0.042	0.162	0.296 ***	0.269 ***	0.333 ***	0.142	0.151
Participant Ability	0.00	-0.057	0.146	0.301 ***	0.262 ***	0.328 ***	0.125	0.124

Significance is denoted as *** = $p < .001$, (corrected for multiple comparisons using Holm's (1979) procedure).

4.4.3 Discussion

The calibration study resulted in an explanatory item response model (i.e. a mixed effects model) that explains performance on the PIAT through four variables of musical structure. As found in the exploratory study, task difficulty increased with the number of imagined arrows (Level), regardless of whether the probe matched the correctly imagined note or not. However, the variables capturing the heard range of notes in the set-up period, and the probability of the probe given the last note heard, differ in their function for trials with correct and incorrect probes, which is indicative of a perceptual bias towards higher probability probe tones and an association of large heard range with the "match" response. Incorrect trials with a probe that was 1 step away, rather than 2 steps away, from the correct imagined note also contributed to item difficulty.

This explanatory model therefore defines ability on the PIAT as the ability to maintain and manipulate tones in mental imagery as well as to resist perceptual biases that can lead to incorrect responses. In this respect the model is in line with recent approaches (Thomas et al., 2018) that combine item response theory and signal detection

theory (SDT). In standard SDT, test performance is defined as a measure of participant ability and response bias, with the purpose being to remove response bias, to obtain a more accurate measure of true ability (Thomas et al., 2018). However, our explanatory model incorporates perceptual biases rather than eliminates it, by defining ability on the PIAT as ability to resist perceptual biases and to perform the pitch imagery task correctly. This incorporation of perceptual bias is particularly relevant to music cognition, as going against and playing with perceptual biases, and expectation, is part of active and passive musical behaviour (Aydogan et al., 2018; Herrmann, Henry, Haegens, & Obleser, 2016). Indeed, the knowledge and use of perceptual bias may be linked to musical creativity, a notion consistent with definitions of musical creativity that involve statistical learning (Pearce et al., 2010; Wiggins & Forth, 2015).

The model has an acceptable prediction accuracy and is plausible in terms of the suggested cognitive process model of the PIAT. In addition, model-based ability estimates along with sum scores from the test correlate significantly with self-reported musical training and perceptual abilities. However, performance on the PIAT is not associated with age nor gender and hence the PIAT represents a fair test with respect to these two variables. The explanatory model can therefore serve as the basis for a future computerised automated version of the PIAT.

4.5 General Discussion

The explanatory model features two variables that suggest perceptual biases in the task: the probability of the probe given the last heard note and the range of notes heard in the set-up component. Cognitive or perceptual biases in music perception have not been studied systematically. While studies have shown visuo-spatial biases in pitch perception (Connell, Cai, & Holler, 2013) and perceptual biases in time perception towards regular rhythmic grouping and intensity (Penel & Drake, 2004), the role of perceptual bias is

often difficult to detangle from a given task. Hence this improved PIAT provides a good opportunity to demonstrate and quantify the effect of these perceptual biases in future validations of the task.

The model contributes to our understanding of Edwin Gordon's concept of 'audiation' (Gordon, 1985, 1989b, 1999). Audiation is "the hearing of music in one's mind when the sound is not physically present" (p. 34, Gordon, 1985). The definition then is synonymous with "musical imagery" (Zatorre et al., 2010), yet to Gordon, audiation was a broader concept that encompassed the processes involved in understanding music that has just been heard, recalling music, composing as well as performing (Gordon, 1989b). Gordon theorised that audiation is the central mental faculty that represents musical aptitude, and hence designed tests for measures of audiation for all ages of development from pre-schoolers to adults (Gordon, 1989a). Today these tests continue to be used by music researchers (Puschmann, 2013; Schleuter, 1993), although most recently some have argued that the norms for children and different age groups have not been updated for three to four decades and may no longer be valid (Ireland, Parker, Foster, & Penhune, 2018). These batteries used simple same-different tests (i.e. hear a melody, insert pause, hear another melody and have participants indicate if second melody was the same or different as the first), where patterns differed either in pitch or rhythm (Gordon, 1989a). While Gordon does not provide a cognitive model of the processes underlying the performance on his tests, recently cognitive models of melodic discrimination tests have pointed to memory and similarity comparison as two core components (P. M. C. Harrison et al., 2016). Yet these same-difference tests cannot be simply equated to musical aptitude or melodic memory abilities as they draw on a number of distinct cognitive processes which contribute to individual differences (P. M. C. Harrison et al., 2016). In addition, Gordon's tests do not require the internal mental

manipulation of sounds or musical elements which is a core component of his audition concept (Gordon, 1989b). In contrast, the PIAT explicitly requires internal manipulation (as well as memory and similarity comparison) as part of the cognitive process for solving the task, making it a better test of Gordon's audiation theory. The results of the current studies show a positive association between self-reported musical training as well as perceptual abilities, and ability on the PIAT. Hence future work will use the PIAT longitudinally to assess ability as children are developing their musical skills, to determine whether this ability to maintain and manipulate tones does in fact predict musical aptitude.

In conclusion, ability on the PIAT requires the skill to both maintain and manipulate tones in mental imagery, as well as to resist perceptual biases that can lead to incorrect responses. Future studies will seek to validate this explanatory model through computerised automated testing, which adapts the difficulty of subsequent items based on updated estimates of the participant's ability on the task. Such validations will also compare performance on established measures of melodic discrimination (P. M. C. Harrison et al., 2016), visuo-spatial working memory (Vock & Holling, 2008), auditory working memory tasks such as the backwards digit span (Wechsler, 2008).

4.6 References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*(12), 1664-1668. doi:10.1016/S0028-3932(00)00079-8

Aydogan, G., Flaig, N., Ravi, S. N., Large, E. W., McClure, S. M., & Margulis, E. H. (2018). Overcoming bias: Cognitive control reduces susceptibility to framing effects in evaluating musical performance. *Scientific Reports, 8*(1), 6229. doi:10.1038/s41598-018-24528-3

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32. doi:10.1023/A:1010933404324

Cebrian, A. N., & Janata, P. (2010). Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Research, 1342*, 39-54. doi:10.1016/j.brainres.2010.04.026

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition, 81*(1), 124-130. doi:10.1016/j.bandc.2012.09.005

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software, 39*(12). doi:10.18637/jss.v039.i12

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.

Deutsch, D. (1970). Dislocation of tones in a musical sequence: a memory illusion. *Nature, 226*, 286. doi:10.1038/226286a0

Deutsch, D. (1972). Effect of repetition of standard and of comparison tones on recognition memory for pitch. *Journal of Experimental Psychology, 93*(1), 156-162. doi:10.1037/h0032496

Gelding, R. W., Thompson, W. F., & Johnson, B. W. (2015). The pitch imagery arrow task: effects of musical training, vividness, and mental control. *PLoS ONE, 10*(3), e0121809. doi:10.1371/journal.pone.0121809

Gordon, E. E. (1985). Research studies in audiation: I. *Bulletin of the Council for Research in Music Education, 84*, 34-50.

Gordon, E. E. (1989a). *Advanced measures of music audiation*. Chicago: Riverside Publishing Company.

Gordon, E. E. (1989b). *Learning Sequences in Music: Skill, Content and Patterns*. Chicago, USA: G.I.A. Publications.

Gordon, E. E. (1999). All about audiation and music aptitudes. *Music Educators Journal, 86*(2), 41-44. doi:10.2307/3399589

Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 1-27). Hillsdale, N.J.: L. Erlbaum Associates.

Halpern, A. R. (2003). Cerebral substrates of musical imagery. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 217-230). New York, NY: Oxford University Press; US.

Halpern, A. R. (2012). Dynamic aspects of musical imagery. *Annals of the New York Academy of Sciences, 1252*(1), 200-205. doi:10.1111/j.1749-6632.2011.06442.x

Halpern, A. R. (2015). Differences in auditory imagery self-reported predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain, 25*(1), 37-47. doi:10.1037/pmu0000081

Halpern, A. R., Golden, H. L., Magdalinou, N., Witoonpanich, P., & Warren, J. D. (2015). Musical tasks targeting preserved and impaired functions in two dementias. *Annals of the New York Academy of Sciences, 1337*(1), 241-248. doi:10.1111/nyas.12616

Harrison, P. M. C. (In Preparation). psychTestR: An R package for creating behavioural experiment interfaces. doi:https://github.com/pmcharrison/psychTestR

Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports, 7*(1), 3618. doi:10.1038/s41598-017-03586-z

Harrison, P. M. C., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: Descriptive and explanatory approaches. *Journal of New Music Research, 45*(3), 265-280. doi:10.1080/09298215.2016.1197953

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience, 24*(6), 1382-1397. doi:10.1162/jocn_a_00216

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience, 28*(11), 2352-2360. doi:10.1111/j.1460-9568.2008.06515.x

Herrmann, B., Henry, M. J., Haegens, S., & Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. *NeuroImage, 124, Part A*, 487-497. doi:10.1016/j.neuroimage.2015.09.019

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Hubbard, T. L. (2018). Some methodological and conceptual considerations in studies of auditory imagery. *Auditory Perception & Cognition*, 1-36. doi:10.1080/25742442.2018.1499001

Ireland, K., Parker, A., Foster, N., & Penhune, V. (2018). Rhythm and melody tasks for school-aged children with and without musical training: Age-equivalent scores and reliability. *Frontiers in Psychology, 9*(426). doi:10.3389/fpsyg.2018.00426

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception & Psychophysics, 68*(5), 829-844. doi:10.3758/BF03193705

Janitza, S., Strobl, C., & Boulesteix, A.-L. (2013). An AUC- based permutation variable importance measure for random forests. *BMC Bioinformatics, 14*, 119. doi:10.1186/1471- 2105-14-119

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences, 1252*(1), 206-213. doi:10.1111/j.1749-6632.2011.06439.x

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception: An Interdisciplinary Journal, 28*(1), 27-46. doi:10.1525/mp.2010.28.1.27

Knoblauch, K. (2014). psyphy: Functions for analyzing psychophysical data in R. doi:http://cran.r-project.org/package=psyphy

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics, 14*(1), 90-94. doi:10.3758/BF03198621

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature, 434*(7030), 158.

Kumar, S., Sedley, W., Barnes, G. R., Teki, S., Friston, K. J., & Griffiths, T. D. (2014). A brain basis for musical hallucinations. *Cortex, 52*, 86-97. doi:10.1016/j.cortex.2013.12.002

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *The Journal of Neuroscience, 29*(8), 2477-2485. doi:10.1523/jneurosci.4921-08.2009

Lee, S. Y., Seok, H., Kim, S.-H., Park, M., & Kim, J. (2018). Immediate effects of mental singing while walking on gait disturbance in hemiplegic stroke patients: A feasibility study. *Annals of Rehabilitation Medicine, 42*(1), 1-7. doi:10.5535/arm.2018.42.1.1

Linden, D. E. J., Thornton, K., Kuswanto, C. N., Johnston, S. J., van de Ven, V., & Jackson, M. C. (2011). The brain's voices: Comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex, 21*(2), 330-337. doi:10.1093/cercor/bhq097

Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Thousand Oaks, C.A.: Sage.

MacKisack, M., Aldworth, S., Macpherson, F., Onians, J., Winlove, C., & Zeman, A. (2016). On picturing a candle: The prehistory of imagery science. *Frontiers in Psychology, 7*, 515. doi:10.3389/fpsyg.2016.00515

Mazerolle, M. J. (2017). AICcmodavg: Model selection and multimodal inference based on (Q)AIC(c) doi:http://cran.r-project.org/package=AICcmodavg

Mullensiefen, D., Ginras, B., Musil, J. J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE, 9*(2), e89642. doi:10.1371/journal.pone.0089642

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50*(1), 302-313. doi:10.1016/j.neuroimage.2009.12.019

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562. doi:10.3758/BF03194900

Puschmann, S. (2013). Pitch-induced responses in the right auditory cortex correlate with musical ability in normal listeners. *NeuroReport, 24*(15), 841-845. doi:10.1097/WNR.0b013e3283650e1e

R Core Team. (2014). R: A language and environment for statistical computing.

Reisberg, D. (1992). *Auditory imagery* (D. Reisberg Ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Schaefer, R. S. (2017). Music in the brain: imagery and memory. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition*. Abingdon: Routledge.

Schleuter, S. L. (1993). The relationship of AMMA scores to sightsinging, dictation, and SAT scores of university music majors. *Contributions to Music Education, 20*, 57-63.

Seashore, C. E. (1919). Musical Imagery and Imagination *The Psychology of Musical Talent*. New York, NY, US: Silver, Burdett & Company. .

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701-703. doi:10.1126/science.171.3972.701

Shinosaki, K., Yamamoto, M., Ukai, S., Kawaguchi, S., Ogawa, A., Ishii, R., . . . Takeda, M. (2003). Desynchronization in the right auditory cortex during

musical hallucinations: A MEG study. *Psychogeriatrics, 3*(2), 88-92. doi:10.1046/j.1479-8301.2003.00009.x

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods, 14*(4), 323–348. doi:10.1037/ a0016973

Thagard, P. (2005). *Mind: Introduction to Cognitive Science* (2nd Edition ed.). Cambridge, Massachusetts: MIT Press.

Thomas, M. L., Brown, G. G., Gur, R. C., Moore, T. M., Patt, V. M., Risbrough, V. B., & Baker, D. G. (2018). A signal detection–item response theory model for evaluating neuropsychological measures. *Journal of Clinical and Experimental Neuropsychology, 40*(8), 745-760. doi:10.1080/13803395.2018.1427699

Vock, M., & Holling, H. (2008). The measurement of visuo–spatial and verbal–numerical working memory: Development of IRT-based scales. *Intelligence, 36*(2), 161-182. doi:10.1016/j.intell.2007.02.004

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th Edition ed.). San Antonio, TX: Pearson.

Weir, G., Williamson, V. J., & Mullensiefen, D. (2015). Increased involuntary musical mental activity is not associated with more accurate voluntary musical imagery. *Psychomusicology: Music, Mind & Brain, 25*(1), 48-57. doi:10.1037/pmu0000076

Wiggins, G. A., & Forth, J. (2015). IDyOT: A Computational Theory of Creativity as Everyday Reasoning from Learned Information. In T. R. Besold, M. Schorlemmer, & A. Smaill (Eds.), *Computational Creativity Research: Towards Creative Machines* (pp. 127-148). Paris: Atlantis Press.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron, 47*(1), 9-12. doi:10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22*(4), 775-789. doi:10.1162/jocn.2009.21239

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*(1), 29-46. doi:10.1162/jocn.1996.8.1.29

Chapter 5

Pitch Imagery Depends Upon Coordination of Auditory and Sensorimotor Brain Activity

Rebecca W. **Gelding**[a,c], William F. **Thompson**[b,c], Blake W. **Johnson**[a,c]

[a] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[b] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

[c] ARC Centre of Excellence in Cognition and its Disorders

Author Contributions

All authors conceived and designed the experiment. R.W.G. conducted the MEG experiment. R.W.G. & B.W.J. analysed the data. All authors discussed the results, wrote and edited the manuscript.

Abstract

Recent magnetoencephalography (MEG) studies have established that sensorimotor brain rhythms are strongly modulated during mental imagery of musical *beat* and *rhythm*, suggesting that motor regions of the brain are important for temporal aspects of musical imagery. The present study examined whether these rhythms also play a role in non-temporal aspects of musical imagery including musical *pitch*. Brain function was measured with MEG from 19 healthy adults while they performed a validated musical pitch imagery task and two non-imagery control tasks with identical temporal characteristics. A 4-dipole source model probed activity in bilateral auditory and sensorimotor cortices. Significantly greater β-band modulation was found during imagery compared to the control tasks of auditory perception and mental arithmetic. Imagery-induced β-modulation showed no significant differences between brain sources, which may reflect a tightly coordinated mode of communication between sensory and sensorimotor cortices via β-oscillations. Directed connectivity analysis revealed that, in the θ-band, the left sensorimotor region drove bilateral auditory regions during imagery onset as well as during perceptual temporal prediction. These results add to the growing evidence that motor regions of the brain are involved in the top-down generation of musical imagery, and that imagery-like processes may be involved in perception.

5.1 Introduction

The silent generation of music in one's own mind is a common experience amongst both musicians and non-musicians. Musical imagery is well developed in musicians: Mozart reportedly experienced his compositions as complete works in his mind through his polyphonic imagery (Agnew, 1922), while Beethoven was (presumably) forced to rely on musical imagery to compose his late symphonies, a period when he was largely or completely deaf (Deutsch & Pierce, 1992). The pianist Glenn Gould had unimpaired hearing but *preferred* to study music by reading it rather than playing it, indicating that musical imagery can be a powerful strategy in and of itself, and not merely a backup strategy necessitated by deafness (Otto, 1990). Among contemporary composers, functional magnetic resonance imaging (fMRI) measurements show that Sting uses highly similar brain regions when listening to or imagining music (Levitin & Grafton, 2016). However, musical imagery is not restricted to the musical elite; the mental replaying of music is a routine, everyday experience for non-musicians, and involuntary musical imagery or "earworms" is common (Bailes, 2015; Farrugia et al., 2015).

Imagery and perception are thought to use overlapping neural mechanisms (Kosslyn & Thompson, 2003), and auditory brain regions are active during musical imagery (Halpern, 2003; Halpern & Zatorre, 1999; Zatorre, 2012). Recent work has also pointed to an important role for motor regions of the brain in auditory perception and imagery. A number of fMRI studies have reported increased BOLD responses in motor areas during musical imagery including primary motor, pre-motor, parietal and inferior frontal cortex (Foster & Zatorre, 2010b; Herholz et al., 2012; Zatorre et al., 2010; Zvyagintsev et al., 2013). Supplementary motor area (SMA) and pre-SMA become active both during motor sequence learning and during anticipation of sound sequences, suggesting the use of motoric predictive mechanisms in both domains (Leaver et al.,

2009). There is also evidence that individual differences in vividness of auditory imagery are correlated with grey matter volume in the SMA, parietal and prefrontal regions, suggesting that the generation of auditory images requires access to auditory-motor representations (Lima et al., 2015). Moore (2011) and Schaefer (2017) have theorised the overlap in imagery and perception activation may reflect a type of imagery ("constructive imagery") that underlies perception, enabling temporal prediction. Taken together, these lines of evidence suggest that music perception and imagery require the coordination of both auditory and motor regions of the brain, even when no overt actions are required. Such findings are consistent with current neurophysiological frameworks that posit ventral and dorsal streams for processing auditory information comparable to those of the visual system (Chen et al., 2008; Rauschecker & Scott, 2009; Zatorre et al., 2007).

Beta-band (β: ~13-30 Hz) oscillations provide a robust neurophysiological marker of motor cortical function that can be measured noninvasively with MEG (Cheyne, 2013; Engel & Fries, 2010). Recent work has implicated this rhythm in temporal predictions of sounds and sound sequences (Arnal, Doelling, & Poeppel, 2015; Fujioka et al., 2015; Fujioka et al., 2012; Iversen et al., 2009; Patel & Iversen, 2014). Beta-band responses to sounds show a characteristic event related desynchronisation (ERD) at a latency of around 200 ms after sound onset, followed by a rebound event related synchronisation (ERS). Fujioka et al. (2012) found that the slope of the β-ERS peaks just before an expected tone. The β-ERD also varies as a function of a listener's metrical interpretation of a simple rhythmic pattern (Iversen et al., 2009). Recent evidence has also confirmed that β-band modulation reflects predictability of pitch (that is, 'what') and not just timing (or 'when' an event will occur), with greater trial-by-trial β-ERD prior to a predictable tone related to reduced P3a amplitude after that tone (Chang et al., 2016, 2018). Finally, physical and imagined accents on a downbeat modulate the β-band response, suggesting that the β-

band plays a role in the temporal coordination of auditory and motor operations in both music perception and imagery (Fujioka et al., 2015; Iversen et al., 2009). Thus, converging evidence suggests that β-band oscillations are functionally associated with the temporal and spectral features of music perception, as well as the temporal features of musical imagery including rhythm, accent and beat.

Recent approaches using magnetoencephalography (MEG) have isolated the primary auditory and sensorimotor regions to investigate how they coordinate in perceptual activities, specifically in beat perception (Tal et al., 2017), temporal prediction (Morillon & Baillet, 2017) and passive listening (Ross et al., 2017). Results from these studies suggest the right auditory cortex is primarily involved in beat perception (Tal et al., 2017). Using directed Phase Transfer Entropy (dPTE), a connectivity analysis for measuring the direction of communication between regions at different frequency bands, Morillon and Baillet (2017) showed that in the β-band (18 – 24 Hz), the bilateral sensorimotor regions drive activity in the auditory cortex during temporal prediction, whilst in lower frequencies (2 – 4 Hz) the auditory regions drive the sensorimotor regions. Enhanced functional connectivity between the left sensorimotor and the bilateral auditory regions has also been found during passive listening to an unfamiliar instrument, after short-term motor training in the action required to play the instrument (Ross et al., 2017). Given the overlap in activation in these areas during perception and imagery, investigating the coordination and connectivity of the primary auditory and sensorimotor regions could prove to be fruitful in the exploration of similarities and differences between musical imagery and perception.

Hence in the present study, we firstly investigated the functional specificity and neurophysiology of β-band oscillations during musical imagery in bilateral auditory and sensorimotor regions of interest. Secondly, we compared the directed connectivity

between the regions during musical imagery and perception. In the imagery task, participants were prompted with visual cues of up and down arrows to silently imagine successive changes in the pitch of a piano note up and down the major scale. Performance on this imagery task was then objectively measured by requiring participants to decide if a probe note did or did not correspond to the result of the specified manipulations. Beta-band oscillations measured during the musical imagery task were compared to those measured during two tasks with identical temporal features but that did not require musical imagery: a music perception task, and a mental arithmetic task.

Our experiment was designed to address three main questions: (1) Does musical imagery differentially modulate β-band oscillations relative to other mental operations (music perception, and mental arithmetic) with identical temporal features? (2) Are β-band oscillations differentially modulated by musical imagery in different brain regions (sensorimotor versus auditory) or hemispheres (right versus left)? (3) Do musical imagery and perception differ in how they direct functional connections between bilateral sensorimotor and auditory regions?

5.2 Materials and Methods

5.2.1 Participants

64 participants were initially trained and screened for imagery performance in a separate behavioural testing session. We selected those who obtained a score of greater than 70% in imagining 3 successive pitch transformations in the Pitch Imagery Arrow Task (PIAT; Gelding et al., 2015), and who reported using a musical imagery strategy to complete the task. 31 participants met the screening criteria and of these 19 participants (14 females, mean age = 25 years; range: 18 – 49 years) were recruited for the MEG study. All participants provided written consent and all procedures were approved by the

Macquarie University Human Research Ethics Committee. Participants ranged in musical training from novice (no formal training) to professional musicians. On average, participants had spent 63% of their life's years actively engaging in musical activities, and 2 out of 19 had negligible musical training.

5.2.2 Stimuli and Apparatus

Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA) was used to control the experiment and to record responses. Acoustic stimuli were generated from the 'Piano' instrument sound in Finale 2012 software (Makemusic Inc; Eden Prairie, MN) and exported as .wav files for use in Presentation®.

5.2.2.1 Pitch Imagery Arrow Task (PIAT)

This task was designed to reliably induce pitch imagery in individuals with a range of musical training (Gelding et al., 2015), and was combined with two control conditions: Perception and Mental Arithmetic (Maths). The Imagery (PIAT) trials began with an ascending major scale to provide a tonal context. A start note (either tonic or dominant of the scale) was then presented simultaneously with the visual presentation of a dot on the screen. A variable number, between 3 – 5, of up / down arrows was next displayed in random order, with each arrow accompanied by a corresponding pitch that moved up / down the scale in a stepwise motion. Pitch changes always matched the direction indicated by the arrows. For the Imagery trials, these stimuli were followed by a continuation phase consisting of 3 silent arrows, in which participants were required to imagine the corresponding pitch steps. A pre-probe screen appeared 1 second before an audible probe pitch, and participants indicated whether the probe matched the final imagined tone. The probe was correct 50% of the time, and when incorrect was within the key signature and a maximum of 2 steps away from correct answer, so as not to make the probe obviously wrong.

Perception trials were identical to Imagery trials but with no continuation component. After the scale and start note, and a series of 3 or more sounded arrows, the pre-probe screen displayed, indicating that participants needed to maintain in memory the last pitch they had just heard. The probe was presented 1 second later, and participants indicated if the probe matched the last tone heard. In this way participants were not aware that they were completing a Perception trial until the pre-probe screen appeared, signalling the end of the sequence of arrows. The probe for Perception trials was set in the same way as for the Imagery trials.

Maths trials began with the instruction "Begin Mental Arithmetic" in silence. A starting number then displayed on the screen, followed by a series of the same visual presentation of arrows, but this time in silence, with numbers above / below the point of the arrows corresponding to addition (up arrow) and subtraction (down arrow) of an ongoing mental calculation. The final probe was a visual number that corresponded to a correct or incorrect answer to this calculation. The complete trial was completed in silence. Accuracy was defined as the percent correct within each condition and mean hit reaction time (ms) was taken as the mean response time from the onset of the auditory probe (Imagery and Perception) or visual number (Maths) until the response was made. For more details about the task see Gelding et al. (2015).

5.2.2.2 MEG Task

The original PIAT was adapted and modified in several respects for the purposes of the MEG study. First, all Imagery trials required 3 imagined tones (rather than variable, ranging from 1-5). Second, conditions all had equal numbers of trials (n = 80 trials) and were randomly interleaved rather than blocked. Finally, no feedback was provided to participants during the task. Figure 5-1 shows a schematic of a PIAT Imagery trial in the MEG study. The last three arrows presented in each condition, either silent (Imagery),

sounded (Perception) or with numbers in silence (Maths) were used in the analysis, as all had a similar visual presentation of arrows at the same rate (one per second).

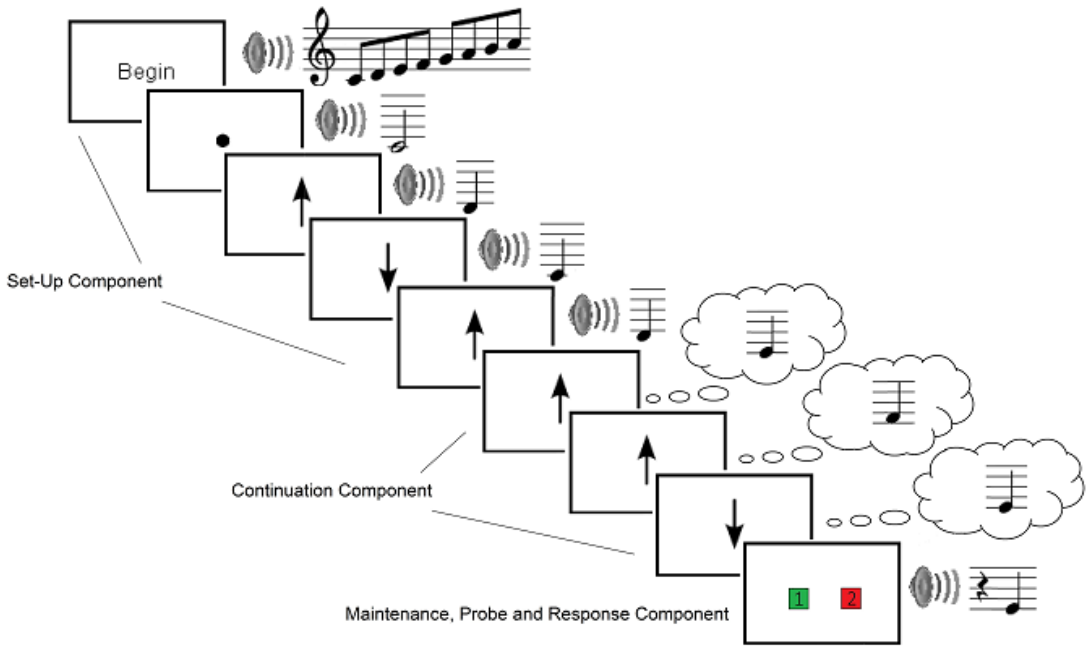

Figure 5-1: Schematic of example MEG PIAT trial. A modified PIAT was used with all Imagery trials set to have 3 imagined tones per trial. Number of heard arrows in set-up component ranged between 3 – 5, Start Note was either the tonic or dominant note of the scale. Key Signature varied randomly between C, C#, D, Eb, E Major. Arrows were presented at a constant rate of 1 per second.

5.2.2.3 Bucknell Auditory Imagery Scale

Participants completed the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015). The 7-point Likert scale includes two subscales, for vividness (BAIS-V) and control (BAIS-C), both of which have 14 items each. BAIS-V questions required participants to rate how clearly they could imagine a particular auditory image. For example, “imagine the sound of a trumpet beginning to play the song Happy Birthday”.

The scale ranged from 1 (“no image present at all”) to 7 (“as vivid as actual sound”), with 4 indicating “fairly vivid”. The BAIS-C questions required participants to rate how easily they could change or manipulate these images at will. For example, “imagine the sound of a trumpet beginning to play the song Happy Birthday. The trumpet stops, and a violin continues the piece.” Again, the scale ranged from 1 (“no image present at all”) to 7 (“extremely easy to change the image”) with 4 indicating “could change the image, but with effort”. The scale items have high reliability (BAIS-V: $\alpha = .83$, BAIS-C: $\alpha = .81$, total scale: $\alpha = .91$) (Halpern, 2015).

5.2.3 Data Acquisition and Analysis

5.2.3.1 Data Acquisition

Brain activity was recorded with a whole-head MEG system (Model PQ1160R-N2, KIT, Kanazawa, Japan) consisting of 160 coaxial first-order gradiometers with a 50 mm baseline. Prior to MEG measurements, five marker coils were placed on the participant's head and their positions and the participant's head shape were measured with a pen digitiser (Polhemus Fastrack, Colchester, VT). Head position was measured by energizing the marker coils in the MEG immediately before and after each block within the recording session. During acquisition MEG was sampled at 1 kHz and band-pass filtered between 0.03 and 200 Hz. Individual structural magnetic resonance images were not available for the present experiment so the adult template brain in BESA Research 6.1 (BESA Research, Gräfelfing, Germany) was used for all participants, using a spherical head model.

There were 80 trials in each of the three conditions. Approximately every 15 trials participants received rest time consisting of a blank screen for several seconds. Participants were asked to remain as still as possible and to limit eye blinks to the inter

trial period or beginning of each trial. Responses were made with the right index finger using a MEG-compatible 2 x 2 button box (Current Designs Inc: Philadelphia, USA Model: HHSC-2x2). Six practice trials were administered inside the MEG, two of each condition type, in which the participant received feedback if they were incorrect. The average time taken to complete the task in the MEG was approximately 56 minutes.

Surface electromyography (EMG) using BrainAmp ExG MR 16P (BrainProducts Gmbh, Gilching, Germany) was also measured using two pairs of bipolar electrodes attached to the orbicularis oris and laryngeal muscle to rule out any systematic muscle activities during the experiment that would indicate vocalisation of the imagined notes or mental arithmetic.

5.2.3.2 Source Localisation

MEG analyses were carried out in BESA Research 6.1. Our spatial filtering approach used four dipole sources modelled in bilateral auditory cortices and bilateral sensorimotor cortices. Auditory sources were fit to each participant's data using the rising half of the averaged M100 response to the onset of each (sounded) arrow in both Imagery and Perception trials (~400 tones). Locations of sensorimotor sources were obtained from beamformer analysis of the β-band ERD to all button press responses (240 trials) for each individual participant. The auditory and sensorimotor sources were then combined into a single model for each participant and time frequency analysis was conducted for each source.

The source localisation procedure resulted in the reduction of the 160 channel surface MEG data to a 4-source montage and provided spatial filters in close spatial proximity to bilateral Auditory and Sensorimotor cortices (Figure 5-2). Subsequent analyses were computed using the 4-source montage and using data only from trials with correct responses.

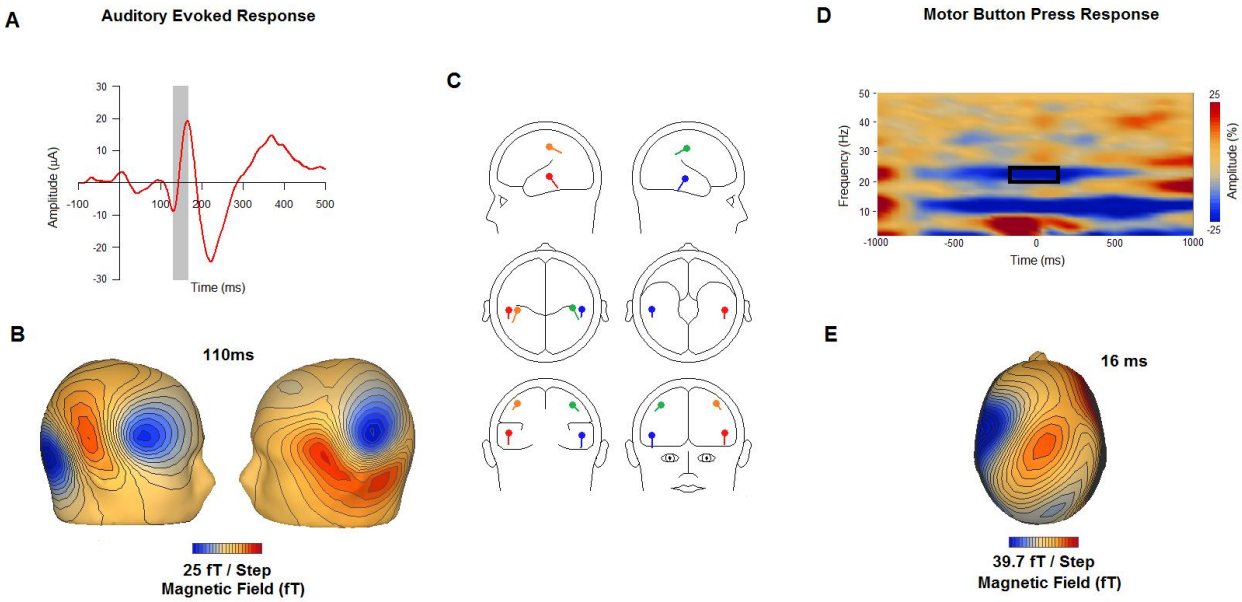

Figure 5-2: A: Left auditory source waveform from single participant (19) of the average sounded arrow across both Perception and Imagery conditions (n = 422) with shaded region indicating time window used to fit the auditory dipole. **B:** Bilateral topographical plot of average sounded arrow across both Perception and Imagery conditions for single participant (1) at t = 110 ms. **C:** Mean auditory and sensorimotor sources from all participants. **D:** Time frequency plot for single participant (16) from central sensor, surrounding the average right-hand button press (t = 0 ms) for all conditions. For this participant, bilateral sensorimotor regions were identified through beamformer taken from 19 – 24 Hz; -150 to 150 ms. **E:** Topographic plot of single participant (1) at time of 16 ms after button press from Maths trials, made in silence, (n = 80).

Average Talairach coordinates of the auditory sources in the BESA Research 6.1 standard template brain were x = 52.3 (right), y = -21.1 (posterior), and z = -1.6 (superior) in the right hemisphere (Right Superior Temporal Gyrus, BA = 21) and x = -52.1, y = -23.3, and z = 2.3 in the left hemisphere (Left Superior Temporal Gyrus, BA = 22). To illustrate the presence of auditory sources, Figure 5-2A shows the left auditory evoked response calculated by averaging all sounded arrows across the Imagery and Perception

conditions, with the shaded region highlighting the time window used for this participant to calculate their dipole. Figure 5-2B also shows the bilateral topographical plots for another example participant at 110 ms after the sounded arrows, indicating the presence of bilateral auditory sources.

To localise the sensorimotor sources for each participant, a beamformer analysis in BESA Research 6.1 was computed in the β-band (15 – 30 Hz) for 300 ms around the onset of the button presses on all trials (n = 240 trials). Bilateral sources within the sensorimotor cortices were found for each participant. The average Talairach coordinates of the BESA Research 6.1 standard template brain for these were: x = 38.5 (right), y = -25.4 (posterior), and z = 45.4 (superior) in the right hemisphere (Right Postcentral Gyrus, BA = 3) and x = -40, y = -29.4, and z = 47.6 in the left hemisphere (Left Postcentral Gyrus, BA = 40). Figure 5-2D illustrates the time frequency plot from a central sensor for a single participant, revealing the presence of the β-band ERD surrounding the average button press response. Figure 5-2E illustrates for a single participant a clear sensorimotor source at the button press for all Maths trials (n = 80) that occur in silence. The location of the group mean coordinates in bilateral superior temporal and postcentral regions is shown in Figure 5-2C.

A 5 second analysis epoch was defined with t = 0 s aligning with the onset of the first of the last three arrows in the trial. Hence the epoch consisted of: the last three arrow presentations (0 – 3 s), the retention period (3 – 4 s) and the probe and response period (4 - 5 s). Since there were a variable number of arrows in the set-up component (each trial randomly set between 0 – 2 arrows for Perception or Maths; 3 – 5 arrows for the Imagery trials), the baseline for each trial was defined as the last 900 ms of the 1000 ms inter-trial interval preceding the trial. Trials with MEG artefacts including blinks and eye-movements during 0 – 5000 ms were rejected from the time-frequency calculations using

the artefact scan tool in BESA Research 6.1, which rejects trials based on abnormally high amplitudes (> 5000 μV) or abrupt rises or falls in amplitude (gradients > 2500).

Time-frequency plots were generated in BESA Research 6.1 using Temporal-Spectral Evolution with a frequency range of 2 to 80 Hz, a frequency sampling of 1 Hz and a time sampling of 50 ms. The plots show the amplitude for each time point normalised to the mean amplitude of the baseline epoch for that frequency. A value of the time-frequency plot describes the spectral change of activity at time *t*, relative to the activity during the baseline epoch. A value of +100% means the amplitude is twice as high as during the baseline epoch. The evoked (averaged) signal was subtracted from all trials prior to computing the mean time-frequency transform. Average time courses for the evoked response (3 – 10 Hz), β-band (15 – 25 Hz), and mu (μ) band (8 – 12 Hz) were calculated in MATLAB 8.2 (MathWorks Inc, MA, United States). The μ-band is another oscillation that has been strongly associated with the motor system (Cheyne, 2013). The β-band of 15 – 25 Hz was chosen based on Fujioka et al. (2015). To illustrate the effectiveness of the dipoles, the mean evoked response time courses of the Perception trials for all participants, from auditory and sensorimotor sources were plotted (Figure 5-3). This shows the clear bilateral auditory evoked response at the onset of each arrow / tone presentation (0, 1 and 2 s) as well as the onset of the audible probe (4 s). The left sensorimotor source shows the typical readiness field, beginning from the pre-probe screen at 3 s, in anticipation of hearing the probe and making a response with the right

index finger. Hence this confirms that the auditory and sensorimotor sources were maximally sensitive to the activities expected in those brain regions.

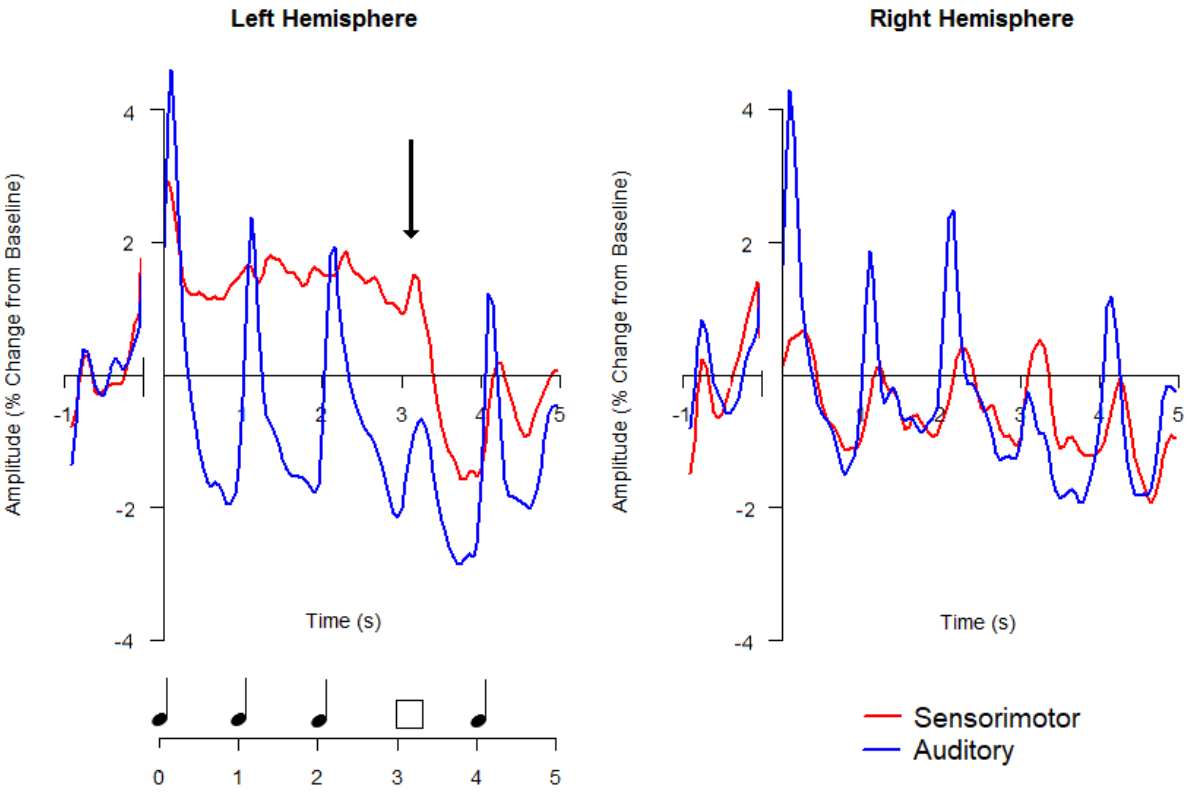

Figure 5-3: 3 – 10 Hz frequency band activity (reflecting the evoked response) from bilateral auditory and sensorimotor sources during Perception trials, for all participants. Stimulus timeline indicates audible tones at 0, 1, 2 s corresponding to the last 3 arrows / tones; square at 3 s indicates presentation of pre-probe screen marking the retention period as participants prepare to hear the probe; audible probe at 4 s and corresponding response with the right index finger. Arrow on left hemisphere plot shows start of the readiness field of sensorimotor cortex in response to the pre-probe screen, as participants prepare to hear probe and respond appropriately with right index finger button press.

5.2.3.3 1 Hz Power Modulation

Given the isochronous presentation of the arrows at a rate of 1 per second in all conditions, the 1 Hz power of the resulting time courses, for each condition and source,

during the last three arrows (0 – 2.95 secs), were calculated using a Fast Fourier Transform, that was zero padded with a frequency resolution of 0.2 Hz. ANOVAs then compared the 1 Hz Power in the three oscillatory bands of interest (Evoked Response, μ-band and β-band) across conditions and sources. To further investigate this modulation, the Maximum ERD and ERS in the β-band and μ-band, relative to the arrow presentation, was calculated per subject. The timing and amount of Maximum ERD was measured at 50 – 450 ms after each of the last three arrow presentations, and averaged per source. Similarly, the timing and amount of Maximum ERS was measured between 500 – 950 ms after each of the last three arrows, and averaged. The difference in both timing and amount of Maximum ERS and ERD were also calculated per arrow, and averaged, measuring the range of modulation as the average time from maximum desynchronisation to reach maximum rebound, as well as the range of amplitude change over this time. ANOVAs were then conducted (3 Conditions (Imagery; Perception; Maths) x 2 Sources (Auditory; Motor) x 2 Hemispheres (Right; Left)) for timing and amount of Maximum ERD and ERS, as well as the difference in timing and amount between Maximum ERD and Maximum ERS, separately for β-band and μ-band. Whenever Mauchley's test of sphericity was significant a Greenhouse-Geisser correction was made. Multiple comparisons were also controlled for using False Discovery Rate (Benjamini & Hochberg, 1995).

5.2.3.4 Directed Phase Transfer Entropy Connectivity

Directed phase transfer entropy (dPTE) is a method that has recently been applied to measuring the effective connectivity between regions of interest (Hillebrand et al., 2016; Lobier, Siebenhuhner, Palva, & Palva, 2014). This method calculates the instantaneous phase of the time series from each region of interest, and, like Granger causality, determines the direction of information based on temporal precedence and

influence of one region on another (Lobier et al., 2014). The advantage of this type of connectivity analysis is that it is model free, computationally straightforward and robust for various time windows and trial numbers (Lobier et al., 2014). To calculate dPTE values, for each participant the raw data time course for each of the last 3 arrows from the four regions of interest, were split into 3 equal epochs across the 1 second interval of an arrow presentation (i.e. 0 – 333 ms, 333 – 666 ms, and 666-999 ms relative to arrow onset). Epochs were concatenated, and band pass filtered into three frequency bands of interest θ (4 – 8 Hz), μ (8 – 12 Hz) and β (15 – 25 Hz). Theta was chosen due to evidence of increased phase connectivity in this frequency band between auditory and sensorimotor regions after short-term motor training (Ross et al., 2017). dTPE was calculated using *PhaseTE_MF.m* (Fraschini & Hillebrand, 2017) with a bin size (*h*) of Scott (1992): $h = 3.49\sigma n^{-1/3}$ where σ is the standard deviation of the phase data calculated from the Hilbert transform for the time series, and *n* is the length (number of samples) of the time series. The default time delay was used, as determined by the frequency content of the data. The results were then normalised between -0.5 and 0.5 with the sign indicating the direction of the connectivity, as per Morillon and Baillet (2017). dPTE values for Imagery and Perception for each connection and each of the three epochs were then compared in t-tests (FDR corrected), to identify significant epochs and connections.

5.3 Results

5.3.1 Behavioural Results

5.3.1.1 Behavioural data

Mean accuracy on the Imagery task was 80.2% (SD = 14.9%), significantly lower than performance on Perception (M = 98.4%; SD = 3%) and Maths (M = 91.9%; SD = 8%) conditions ($F_{(2,54)} = 16.18$; $p < .001$; $\eta^2 = .37$), reflecting greater difficulty of the

Imagery task. Mean hit reaction time (ms) was significantly slower for Imagery trials (M = 834 ms, SD = 169 ms) than for Perception (M = 707 ms, SD = 96 ms) or Maths trials (M = 685 ms, SD = 113 ms) ($F_{(2,54)} = 7.36$; $p = .002$; $\eta^2 = .22$).

Accuracy on the Imagery and Maths tasks was significantly correlated with musical experience index (MEI; $r = .50$, $p = .028$; and $r = .47$, $p = .04$, respectively). In contrast to the results of Gelding et al. (2015), accuracy on the Perception task was not significantly related to MEI, due to a ceiling effect in which all but 3 participants obtained 100% accuracy on the Perception trials.

Debrief vividness ratings (a 1 – 5 self-rated score of vividness experienced during task completion) were significantly positively correlated with Imagery accuracy ($r = .71$, $p < .001$), and negatively correlated with Imagery reaction time ($r = -0.69$, $p = .001$) and Perception reaction time ($r = -0.536$, $p = .018$).

5.3.1.2 Bucknell Auditory Imagery Scale

Imagery accuracy (percent correct) was positively correlated with both the vividness (BAIS-V) ($r = .52$, $p = .023$) and control (BAIS-C) ($r = .48$, $p = .037$) subscales of the BAIS, consistent with the results of our previous behavioural study (Gelding et al., 2015). Neither subscale showed a significant correlation with accuracy in the Perception or Maths condition. The BAIS-C subscale showed a significant negative correlation with mean hit reaction times in both the Imagery ($r = -0.54$, $p = .017$) and Perception ($r = -0.56$, $p = .014$) conditions.

Taken together, the behavioural results provide strong support for our contention that participants performed the Imagery task with a pitch imagery strategy: imagery performance was gauged with an objective behavioural measure; performance was significantly positively correlated with an independent measure of individual imagery

ability (BAIS); and all participants confirmed the use of a musical imagery strategy to complete the task (i.e. hearing the sounds in their head or singing them in their head) in post-experimental debriefings.

5.3.2 MEG Results

5.3.2.1 Normalised β-Band Time Course

The mean β-band time course over 0 – 3 seconds, corresponding to the last three arrows presented in each condition, were calculated from the time-frequency plots. These were then normalised and plotted (Figure 5-4). The absolute area under the (rectified) curve from this figure showed greater mean β-modulation for Imagery (1.168%) than Perception (0.89%) or Maths (0.843%).

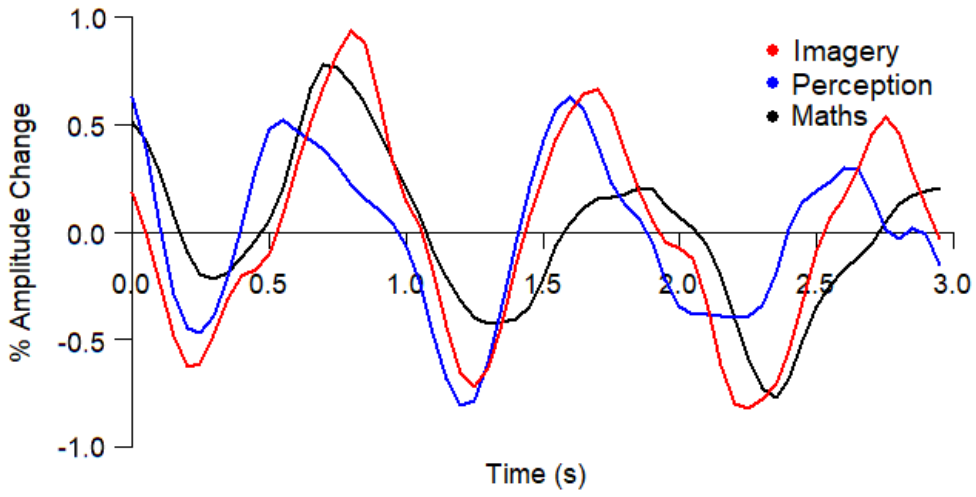

Figure 5-4: Mean normalised β-band time course for Imagery, Perception and Maths Condition during the last 3 arrows of each trial (arrow onset at t = 0, 1, 2 s).

5.3.2.2 Power of 1 Hz Modulation

Evoked Response (3 – 10 Hz)

Repeated measures ANOVA (3 Conditions (Imagery; Perception; Maths) x 2 Sources (Auditory; Sensorimotor) x 2 Hemispheres (Right; Left)) computed for the 1 Hz power of the evoked response time course, over the time window of 0 – 2.95 s were calculated. Given Mauchly's test of sphericity was significant ($p < .005$), Greenhouse-Geisser corrections were made. No significant main effects or interactions were found. Figure 5-5A confirms substantially greater 1 Hz power in the Auditory sources in the Perception condition relative to the Imagery and Maths conditions, however correcting for sphericity revealed these differences were not significant.

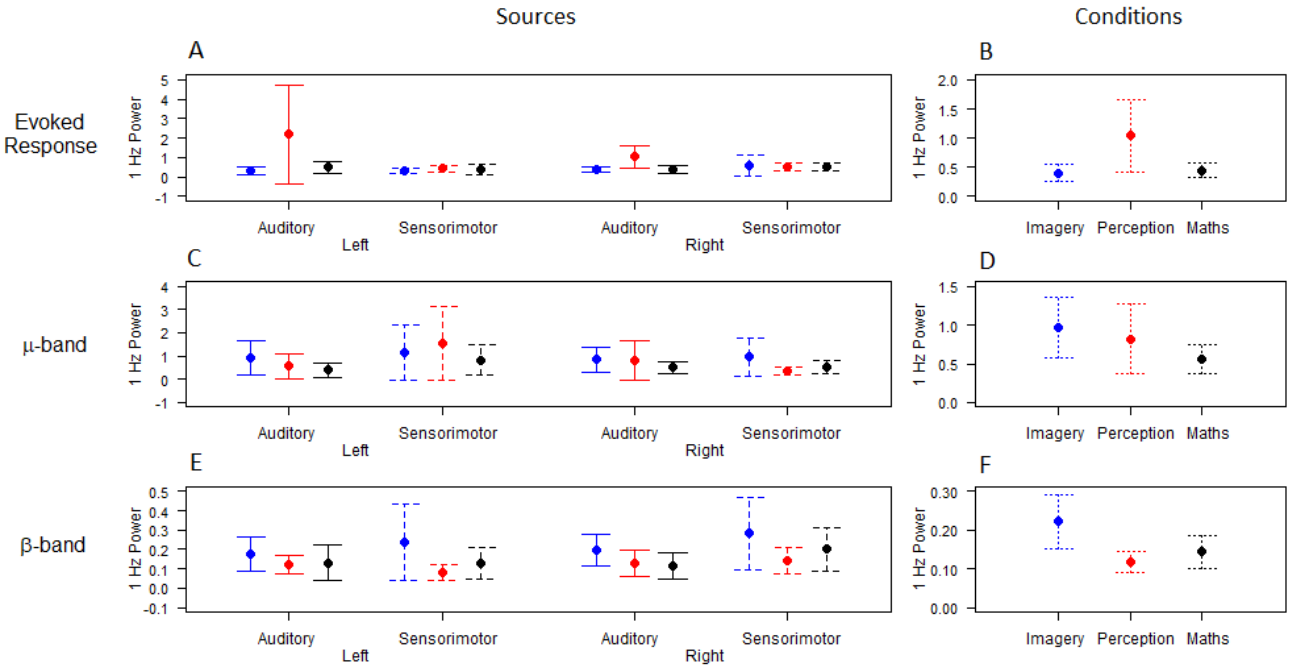

Figure 5-5: 1 Hz Power during continuation period for all three conditions. **A, B:** Evoked Response (3- 10 Hz); **C, D:** μ-band (8 – 12 Hz) and **E, F:** β-band (15 – 25 Hz). 1st Column shows mean results and 95% Confidence Intervals for each of the four sources for the three conditions (Imagery: blue; Perception: red; Maths: black), with auditory sources having solid lines and sensorimotor sources having dashed lines (A, C, E). 2nd Column shows mean results and 95% Confidence Intervals of all sources for each of the three conditions (B, D, F).

μ-Band (8 – 12 Hz)

A 3 x 2 x 2 repeated measures ANOVA for the μ-band 1 Hz power also showed no significant main effects or interactions. This is seen in Figure 5-5D.

β-Band (15 – 25 Hz)

A 3 x 2 x 2 repeated measures ANOVA for the β-band 1 Hz power showed a significant main effect of Condition after correcting for sphericity using Greenhouse-Geisser ($F_{(2,216)} = 4.826$, $p = .0197$, GG $\varepsilon = 0.839$). Post hoc comparisons (Tukey HSD, $\alpha = 0.05$) showed that Imagery power was significantly greater than Perception ($p = .011$), but there was no significant difference between Perception and Maths ($p = .735$) or Imagery and Maths ($p = .076$). This confirms that β-band magnitude was modulated significantly more during pitch imagery than during pitch perception, as seen in Figure 5-5F.

5.3.2.3 Exploratory analyses

β-Band

We carried out additional analysis to explore the sensitivity and power of additional metrics of β-power modulation: timing, and magnitude of Maximum β-ERD (Max ERD_{TIME} and Max ERD_{AMP}, respectively) and Maximum β-ERS (Max ERS_{TIME} and Max ERS_{AMP}, respectively), and the difference between them (Time Difference and Magnitude Difference). All ANOVAs were corrected for Sphericity – where Mauchly's test was significant – as well as False Discovery Rate. Figure 5-6 provides a schematic of the various metrics calculated.

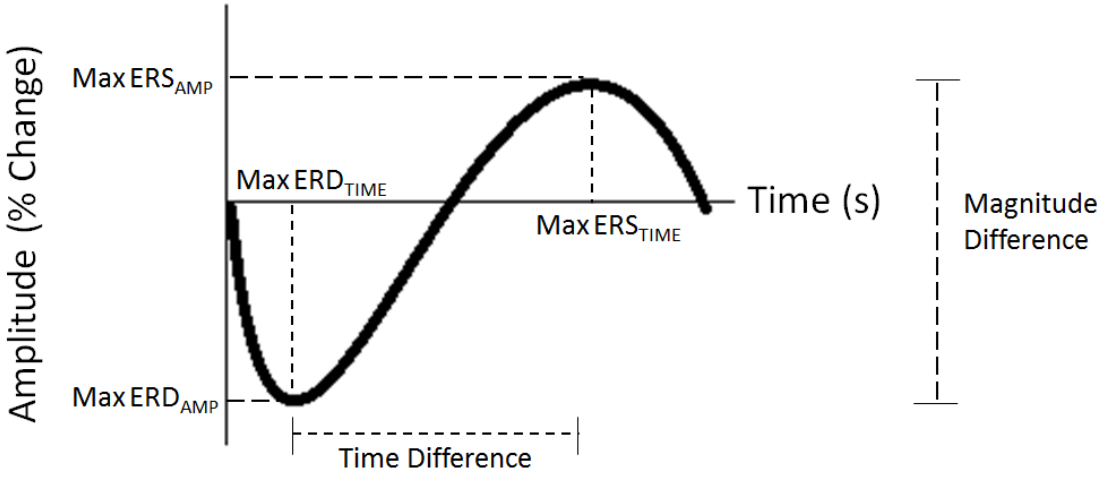

Figure 5-6: Schematic of metrics calculated for exploratory analysis for each 1 second time interval during the continuation period.

Timing of β-ERD

A significant main effect of Condition was found for Max ERD$_{TIME}$ ($F_{(2,216)}$ = 20.26, p = .0001, GG ε = 0.866, FDR corrected). Post hoc Tukey HSD comparisons revealed that the Max ERD$_{TIME}$ in Perception (M = 225.23 ms, SD = 61.19 ms) occurred significantly earlier than both Imagery (M = 262.06 ms, SD = 71.68 ms) (p = .003) and Maths (M = 310.96 ms, SD = 71.42 ms) (p < .0001), and Imagery was earlier than Maths (p < .0001).

Timing of β-ERS

A significant main effect of Condition was also found for Max ERS$_{TIME}$ ($F_{(2,216)}$ = 12.51, p = .003, GG ε = 0.839, FDR corrected). Post hoc Tukey HSD comparisons showed that Max ERS$_{TIME}$ in Perception (M = 679.83 ms, SD = 83.86 ms) occurred again significantly earlier than both Imagery (M = 736.62 ms, SD = 87.01 ms) (p =.0002) and Maths (M = 766.67 ms, SD = 87.73 ms) (p < .0001), but there was no significant difference in the timing of the rebound for Imagery and Maths (p = .083). Figure 5-7

shows a summary of these main results for Max ERD_{TIME} and Max ERS_{TIME}, with Perception β-ERD and β-ERS occurring earlier than the other conditions.

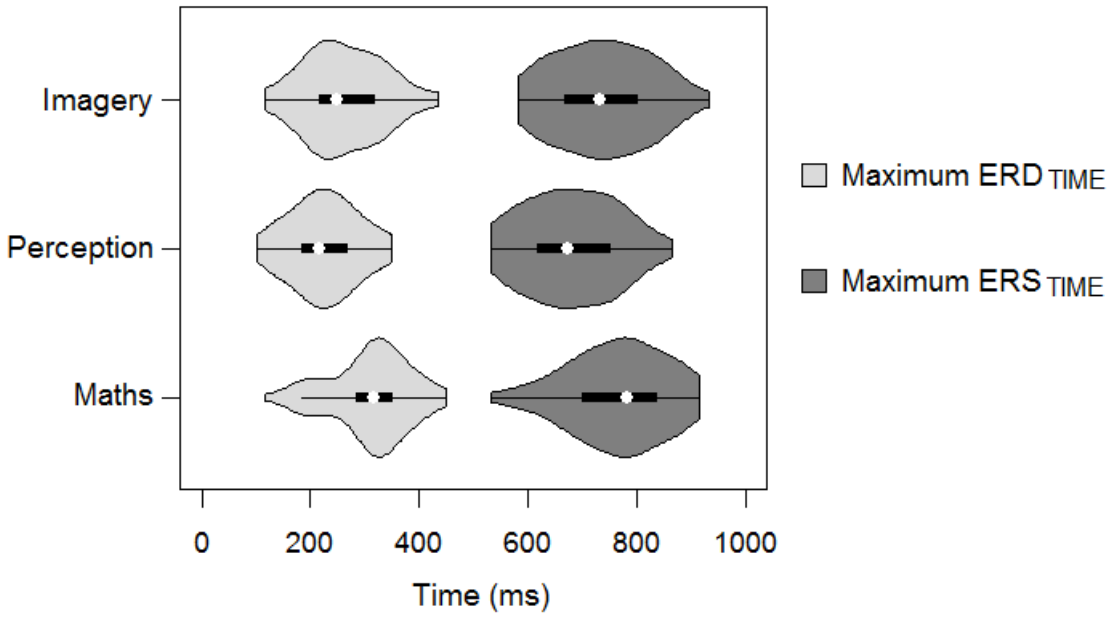

Figure 5-7: β-band: Max ERD_{TIME} (light grey) and Max ERS_{TIME} (dark grey) for all conditions, averaged across all sources. Perception condition was significantly earlier for Max ERD_{TIME} and Max ERS_{TIME} than Imagery or Maths. Imagery condition was faster than Maths for Max ERD_{TIME} but no difference for Max ERS_{TIME}.

Magnitude Difference

A main effect of Condition was also found in Magnitude Difference ($F_{(2,216)}$ = 6.43, p = .0338, FDR corrected). Post hoc Tukey HSD comparisons showed that Imagery (M = 2.6%, SD = 1.2%) had a significantly larger Magnitude Difference (i.e. greater difference between Max ERS_{AMP} and Max ERD_{AMP}) than Perception (M = 2.2%, SD = 0.75%) (p = .029) and Maths (M = 2.2%, SD = 1.1%) (p = .019). There was no difference between Perception and Maths.

These results corroborate the 1 Hz Power analysis and suggest that β-band magnitude was more significantly modulated during Imagery than Perception. Moreover,

given β-band modulation, as measured by the Magnitude Difference metric, was greater for Imagery than Maths, the modulation appears to be specific to pitch imagery, and less evident during silent mathematical computations. No other significant main effects were found in Time Difference or Magnitude Difference.

In addition, to validate the use of the 1 Hz Power FFT as a measure of modulation, the correlation between Magnitude Difference and 1 Hz Power in β-band was calculated, and found to be significant ($r = .811$, $p < .0001$). This finding confirms that both metrics provide similar information about the modulation of oscillations occurring during the last three arrows across conditions.

μ-Band

Although the 1 Hz power ANOVAs in the μ-band did not reveal significant effects, it is still informative to use the same exploratory metrics to compare results with the β-band. The only significant results, after controlling for Sphericity if Mauchly's test was significant, and for False Discovery Rate, were found in the Timing of μ-ERD and μ-ERS.

Timing of μ-ERD

A significant main effect of Condition was found for Max ERD_{TIME} ($F_{(2,216)} = 11.75$, $p = .0028$, FDR corrected). Post hoc Tukey HSD comparisons showed that μ-ERD in Perception (M = 222.59 ms, SD = 91.72 ms) was significantly earlier than Maths (M = 287.06 ms, SD = 100.69 ms) ($p < .0001$), but not significantly different to Imagery (M = 254.17 ms, SD = 84.81 ms) ($p = .095$). There was no significant difference between Imagery and Maths ($p = .078$).

Timing of μ-ERS

A significant main effect of Condition was found for Max ERS_{TIME} ($F_{(2,216)}$ = 8.557, p = .011, FDR corrected). Post hoc Tukey HSD comparisons revealed that μ-ERS in Perception (M = 689.91 ms, SD = 117.22 ms) was significantly earlier than both Imagery (M = 738.82 ms, SD = 104.33 ms) (p = .013) and Maths (M = 768.20 ms, SD = 90.03 ms) (p < .0001). There was no significant difference between Imagery and Maths in the timing of the μ-ERS (p = .198).

5.3.2.4 Connectivity Analysis

The comparison of Imagery and Perception dPTE values revealed several significant differences after FDR correction for the θ-band only. The mean and standard error of the dPTE values for Imagery and Perception are plotted for the right auditory source connections (Figure 5-8A,B), the left auditory source connections (Figure 5-8C,D) and for the hemisphere connections across same sources (Figure 5-8E,F). At the onset of the arrow and tone in the Perception condition (0 – 333 ms epoch), in the θ-band, the bilateral auditory regions (rAUD, lAUD) lead the sensorimotor regions (rSM, lSM) more than during the Imagery condition where the results suggested the opposite direction of information flow (rAUD to lSM: t(18) = -3.575, p = .0194; rAUD to rSM: t(18) = -3.175, p =.0315; lAUD to lSM: t(18) = -3.772, p =.0195, FDR corrected). Interestingly, this pattern is reversed during the last epoch (666 – 999 ms), for the left sensorimotor to right auditory connection, where the left sensorimotor region leads the right auditory in the Perception condition, but the right auditory leads the left sensorimotor in the Imagery condition. However, this difference between conditions was not significant after FDR correction (p = 0.051).

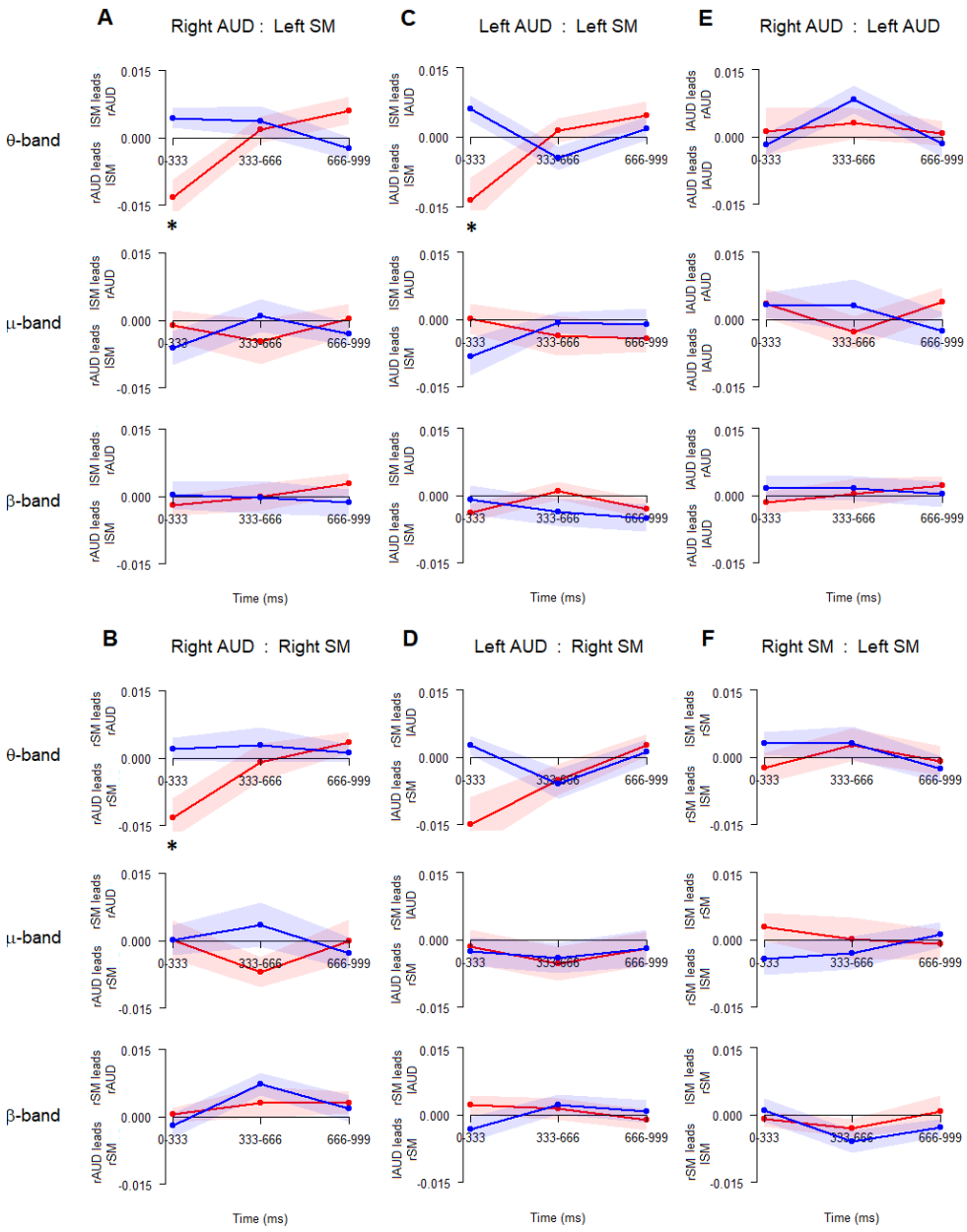

Figure 5-8: dPTE mean values for the three frequency bands of interest, θ (1st, 4th row), μ (2nd, 5th row), and β (3rd, 6th row) during Perception (red) and Imagery (blue) for all connections (via column): **A)** rAUD to lSM; **B)** rAUD to rSM; **C)** lAUD to lSM; **D)** lAUD to rSM; **E)** rAUD to lAUD; **F)** rSM to lSM. Values were calculated over three equal epoch divisions of the 1 second arrow presentation. Shaded regions are the standard error of the mean. (A – D): positive values indicate the sensorimotor region activity leading the auditory region (top-down), negative values indicate the auditory region activity leading the sensorimotor region (bottom-up). (E, F): positive values indicate the left source activity leading the right source, negative values indicate the right source activity leading the left source. * = $p < .05$ FDR corrected.

During the middle epoch (333 – 666 ms), when the image of the tone was being updated, following the visual perception of the arrow, the left auditory region leads the θ-band connectivity to all other regions (Figure 5-8 C,E (1st row); D (4th row)), whilst the left sensorimotor regions also leads both the right auditory region (Figure 5-8 A (1st row) and the right motor region (Figure 5-8 F(4th row)). No significant difference between Imagery and Perception were found in the dPTE in μ-band or β-band between any connections.

Taken together these results show that in the Perception condition, information flow in the θ-band is directed from the auditory to the sensorimotor regions at the onset of sound; however, the (left) sensorimotor cortex drives the bilateral auditory regions during the anticipation of a sound. The latter pattern, is also seen at the onset of silent arrows for the Imagery condition, suggesting the invocation of motorically-driven predictive mechanisms. Further, the left auditory source driving the bilateral sensorimotor and the right auditory activity, along with the left sensorimotor source also driving both right hemisphere sources, during the middle epoch of Imagery, supports the notion of the left hemisphere playing an important role in pitch memory (Schaal et al., 2017; Schaal, Williamson, et al., 2015) and in pitch manipulation (Albouy et al., 2017).

5.3.2.5 Regression Modelling

The above results show that in the θ-band, the left sensorimotor cortex leads the right auditory region in the last epoch in the Perception condition. Since it has been posited that temporal prediction in perception may require imagery (Schaefer, 2017), we explored the nature of this relationship further through regression modelling to determine if imagery ability is facilitated by sensorimotor processing that precedes auditory processing. Linear regression was calculated using the dPTE value of the right auditory to the left sensorimotor connection in Perception condition in the last epoch (666 – 999

ms) as the criterion variable. The predictor variable in the initial regression was Imagery accuracy (percent correct Imagery condition) alone. The result was significant ($F_{(1, 17)}$ = 6.951, p = .017), with Imagery accuracy alone accounting for 29% of the variance in the dPTE value (R^2 = .290; R^2_{adj} = .248) (see Figure 5-9). The addition of BAIS total scores or Imagery reaction time made no significant improvements to the model.

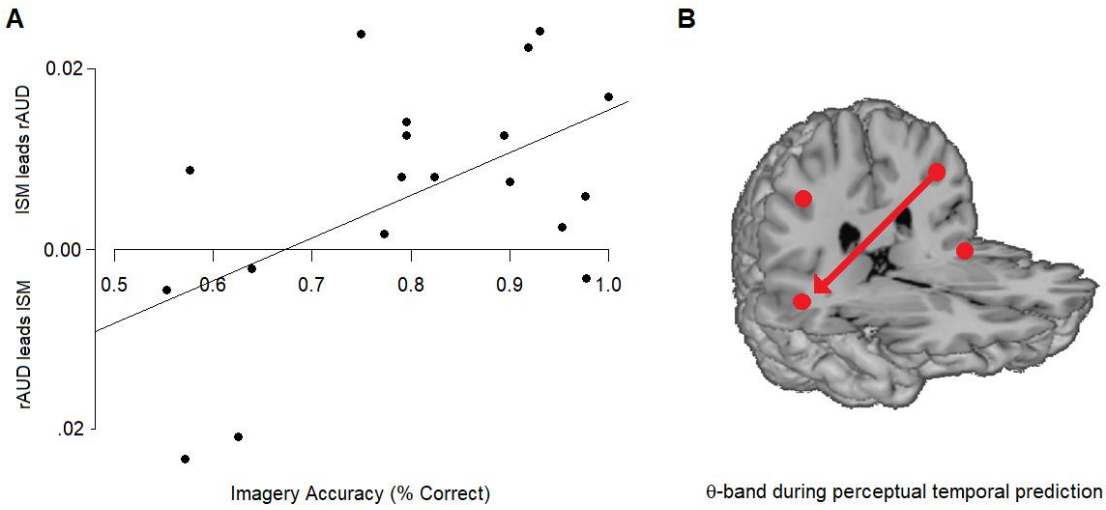

Figure 5-9: A: dPTE for the right auditory to the left sensorimotor connection in the Perception condition, in the time window preceding to the next arrow (666 – 999 ms). **B:** Schematic of directed connectivity between the left sensorimotor and the right auditory region during this period. Imagery accuracy is positively associated with greater top-down directed connectivity from the left sensorimotor region to the right auditory region

5.4 Discussion

The results of this investigation show that greater β-band modulation occurred during pitch imagery than pitch perception or maths, and that connectivity within the θ-band was directed from sensorimotor to auditory regions during imagery generation as well as during perceptual temporal prediction. Whilst the β-band has previously been shown to be involved in predictive timing of 'when' events will occur in both perception (Fujioka et al., 2012) and imagery (Fujioka et al., 2015), only recently has evidence

emerged that it is also important in perception for predicting 'what' events they may be (Chang et al., 2016, 2018). For the first time, our study shows that even in silence, while *imagining* pitches, increased β-band modulation is observed, suggesting that β-band is also involved in 'what' is being imagined. Contrary to our expectations, our analyses showed no significant hemispheric or regional differences in β-band modulation for any type of cognitive operation. There was also no evidence for a differential connectivity during perception and imagery in the β-band. However, in the θ-band different patterns of directed connectivity were seen between perception and imagery. Our results may have implications for our understanding of how sensorimotor brain regions interact to support musical imagery, and how imagery-like processes may support perception.

In any study of a subjective phenomenon such as mental imagery, it is necessary to ensure that the activity of interest is taking place. Our results support the notion that participants were using pitch imagery to complete the task. Firstly, objective behavioural measures of accuracy and reaction time gauged performance and were significantly correlated with a subjective measure of individual imagery ability (BAIS). Secondly, all participants confirmed in post-experimental debriefings the use of a musical imagery strategy to complete the task (i.e. hearing the sounds in their head or singing them in their head).

5.4.1 Modulation of β-Band During Pitch Imagery

Greater β-band power modulation is seen during pitch Imagery than Perception, as evidenced both in the initial normalised time course plot (Figure 5-4) and in the 1 Hz power calculations (Figure 5-5F). The follow up analysis of modulation metrics confirmed that Imagery showed significantly greater Magnitude Difference than both Perception and Maths. This suggests that pitch imagery, completed in silence, not only

significantly modulates the β-band more than perception, but also that this modulation is specific to pitch imagery (and not other silent manipulations such as mental arithmetic).

The dynamics of the β-band in response to sound has been argued to contain both exogenous components, seen in the initial β-ERD, and endogenous components, seen in the subsequent rebound, β-ERS (Fujioka et al., 2012; Ross et al., 2017). Much attention has been given to the latter, suggesting a role of the β-band rebound in predictive timing and rhythm perception over intervals of less than 1 second (Arnal et al., 2015; Arnal & Giraud, 2012; Fujioka et al., 2012), with the same predictive effect not seen in longer inter-stimulus intervals (Meijer, te Woerd, & Praamstra, 2016). In addition, auditory β-band activity tracks the location in time of imagined beats during a presentation of a syncopated rhythm (Snyder & Large, 2005).

In this task the 'when' of arrow presentation was kept constant, while the 'what' varied (i.e. pitch to be imagined or heard, or a mathematical operation). Hence the emphasis during imagery was on pitch manipulations, and not the predictive timing of events. This is due to the arrows providing a visual isochronous metronome at a rate of 1 per second, which minimised the anticipatory timing (Colley, Keller, & Halpern, 2018), and was a slower rate than the human preferred tempo (Patel & Iversen, 2014). Flashing visual metronomes are much less effective for synchronising than auditory metronomes at the same tempos (Iversen, Patel, Nicodemus, & Emmorey, 2015) and do not induce a strong sense of beat (McAuley & Henry, 2010).

Whilst β-band dynamics have been extensively studied in the role of anticipation of auditory beats, recently the role of β-band in predicting not only 'when' but 'what' a pitch will be, have also been explored (Chang et al., 2016, 2018). Chang et al. (2016) used a mismatch negativity paradigm with either 10% or 20% deviant tones and two different presentation rates (500 ms, 610 ms). Discrete Fourier Transforms were calculated for the

resulting β-band time courses and peaks were found at the presentation rate of 2 Hz and 1.6 Hz respectively. They use this as evidence that the β-band entrains to the different inter-onset intervals in isochronous sequences (Chang et al., 2016). Though not compared statistically, the left auditory region 2 Hz power in the 20% deviant appears larger than the 10% deviant condition, suggesting greater entrainment when the deviant is more predictable. This is confirmed in a follow up study, in which Chang et al. (2018) showed that greater β-band modulation was seen prior to a predictable rather than an unpredictable tone in a MMN paradigm. Further, a trial-by-trial analysis revealed that greater β-ERD prior to a predictable deviant tone was related to reduced P3a amplitude after that tone, suggesting a reduced attention-prediction error response (Chang et al., 2018). Our study has shown greater modulation of the β-band during pitch imagery in a task that crucially has a constant presentation rate across all conditions, thereby minimising unexpectedness surrounding timing. This confirms that the β-band is involved not only in 'when' but also 'what' is to occur. Even in the absence of a sound, the internal generation or manipulation of a sound image corresponded to greater modulation of β-band than for perception.

5.4.2 β-band Communication

Our neurophysiological measures were analysed to provide sufficient spatial sensitivity to discriminate broadly between neural activity in the two hemispheres, and between auditory and sensorimotor regions within hemispheres. Within each of the three conditions, similar time courses were seen between the auditory and sensorimotor sources in the β-band. Evidence for this is seen in the 1 Hz power calculation were no differences were found in source or hemisphere in the β-band (Figure 5-5E). In addition, no differences in the timing or amount of Maximum β-ERD or β-ERS were seen between sources or hemispheres. The β-band has historically been viewed as a sensorimotor rhythm (Cheyne, 2013; Engel & Fries, 2010; Kilavik, Zaepffel, Brovelli, MacKay, &

Riehle, 2013). Our results are consistent with the concept of the β-band as an "open line" of communication between the regions (Tang et al., 2016), as predicted by forward and inverse feedback loops on the dorsal pathway (Rauschecker & Scott, 2009). This open line of communication appears to be present during perception, imagery as well as mental arithmetic.

Our behavioural results showed that participants were significantly less accurate in the Imagery task, raising the possibility that the obtained physiological results are due to differences in task difficulty. However, there is a considerable literature showing that mental effort (workload, fatigue and drowsiness) is associated with frequency bands lower than β (i.e. delta, theta and alpha) (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014). Further, μ/α-band power is known to be strongly and monotonically *suppressed* as a function of increased listening effort (Obleser & Weisz, 2012). The present results showed a quite different pattern of results, with an *enhancement* of β-band power during mental imagery and no significant power modulation in lower frequency bands. Condition effects in the μ-band were found only for the timing of μ-ERD and μ-ERS, with faster rebounds in Perception than *both* Imagery and Maths, as would be expected since time locking should be tighter in the Perception condition. This profile of physiological responses is not readily attributable to differences in the difficulty of our tasks.

5.4.3 Directed Connectivity of Auditory and Sensorimotor Cortices

Our connectivity analysis revealed that during the generation of imagery, in the θ-band, the left sensorimotor region is leading the activity in the bilateral auditory regions. Conversely during the onset of a tone/arrow in the Perception condition, this directionality is reversed. Interestingly, the opposite pattern is seen in the last 333 ms of the arrow presentation, with the temporal prediction of the next arrow resulting in greater left sensorimotor to right auditory region direction of connectivity in the Perception

condition, whilst the Imagery condition showed more right auditory to left sensorimotor region directed connectivity. The regression finding that Imagery accuracy significantly predicted 29% of the variance in connectivity during perceptual temporal prediction suggests there is some overlap between temporal prediction and imagery. This has been theorised previously by Moore (2011) and Schaefer (2017) as two different types of imagery, with perception involving a "constructive imagery" and voluntary musical imagery, as is required in the PIAT, involving "sensory imagery". To our knowledge, this is the first empirical evidence for this proposition.

In addition, during the middle epoch of each silent arrow in the Imagery Condition, a left hemisphere dominance was seen with the left auditory region leading the θ-band connectivity to all other regions and the left sensorimotor region also leading the right auditory and sensorimotor regions. This suggests the left hemisphere plays an important role during pitch imagery. Whilst the right auditory regions have repeatedly been found to be involved in musical imagery (Herholz et al., 2012; Hubbard, 2013; Zatorre & Halpern, 2005), the majority of the paradigms used have not isolated pitch imagery from other aspects of the musical image (ie rhythm, melody). In contrast, our results are consistent with the brain stimulation studies of Schaal and colleagues that show the left supramarginal gyrus, but not the right, having a causal role in *pitch* memory (Schaal et al., 2017; Schaal et al., 2013; Schaal, Williamson, et al., 2015). Further, stimulation to the left intraparietal sulcus at the θ-band frequency has been shown to boost performance on manipulation of melodies (i.e. mental reversal) but no improvement after stimulation is seen for maintenance of melodies (Albouy et al., 2017). Hence the left hemisphere dominance seen in the present study may also be due to the manipulation required in the task.

Finally, whilst Morillon and Baillet (2017) found β-band top-down directed connectivity during temporal prediction of a melody, and such connectivity was also expected to be found in during imagery (Gelding & Sun, 2018), we found no evidence for a significant difference in connectivity between perception and imagery in the β-band. For all connections, the normalised dPTE values in both conditions remained around 0. However, our finding that θ-band top-down directed connectivity during temporal prediction was related to imagery performance does confirm the role of the sensorimotor region in both perceptual temporal prediction and musical imagery.

5.5 Conclusion

This is the first study to show that the β-band auditory and sensorimotor oscillations are modulated more by pitch imagery manipulation than by perception, and the first to show opposite directed connectivity during perception and imagery onset. Further, our data provide evidence that imagery performance is associated with the *degree* of connectivity directed from sensorimotor to auditory regions during temporal prediction in perception.

Previously in perception studies, the β-band has been shown to be involved in predictive timing of both 'when' events will occur, and 'what' these events will, however up until now, musical imagery studies have only shown the β-band's involvement in temporal prediction. Our study shows that even in silence, while imagining pitches, increased β-band modulation is seen in comparison to perception suggesting the β-band is also important for 'what' imagined events are. In this task, timing or beat is held at a constant rate and pitches – heard or imagined – varied. Future research should investigate if the same increase in modulation, compared to perception, would be seen in rhythm imagery tasks, in which the pitch is held constant, but the rhythmic patterns vary. While the β-band is proposed to track the temporal location of imagined beats in syncopated

auditory rhythms (Iversen et al., 2009), it remains unclear if the β-band will also track imagined sound onsets in a rhythmic pattern, both on and off the beat. Future research could also confirm whether the presence of imagery-like top-down connectivity is present during other temporal prediction tasks and whether this is an association with explicit musical imagery ability. If so, it would not only suggest the presence of a type of imagery occurring implicitly during perception (Schaefer, 2017), but could lead to new avenues of research into the benefits of imagery training on perceptual abilities, and potential imagery interventions for hearing loss.

5.6 References

Agnew, M. (1922). The auditory imagery of great composers. *Psychological Monographs: General and Applied, 31*(1), 279-287.

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron, 94*(1), 193-206. doi:10.1016/j.neuron.2017.03.015

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex, 25*(9), 3077-3085. doi:10.1093/cercor/bhu103

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390-398. doi:10.1016/j.tics.2012.05.003

Bailes, F. (2015). Music in mind? An experience sampling study of what and when, towards an understanding of why. *Psychomusicology: Music, Mind & Brain, 25*(1), 58-68.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44*, 58-75. doi:10.1016/j.neubiorev.2012.10.003

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2016). Unpredicted pitch modulates beta oscillatory power during rhythmic entrainment to a tone sequence. *Frontiers in Psychology, 7*, 327. doi:10.3389/fpsyg.2016.00327

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2018). Beta oscillatory power modulation reflects the predictability of pitch change. *Cortex, 106*. doi:10.1016/j.cortex.2018.06.008

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex, 18*(12), 2844-2854. doi:10.1093/cercor/bhn042

Cheyne, D. O. (2013). MEG studies of sensorimotor rhythms: A review. *Experimental Neurology, 245*, 27-39. doi:10.1016/j.expneurol.2012.08.030

Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronization with expressively timed music. *The Quarterly Journal of Experimental Psychology, 71*(8), 1781 - 1796. doi:10.1080/17470218.2017.1366531

Deutsch, D., & Pierce, J. R. (1992). The climate of auditory imagery and music. In D. Reiseberg (Ed.), *Auditory Imagery* (pp. 237-260). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations — signalling the status quo? *Current Opinion in Neurobiology, 20*(2), 156-165. doi:10.1016/j.conb.2010.02.015

Farrugia, N., Jakubowski, K., Cusack, R., & Stewart, L. (2015). Tunes stuck in your brain: The frequency and affective evaluation of involuntary musical imagery correlate with cortical structure. *Consciousness and Cognition, 35*, 66-77. doi:10.1016/j.concog.2015.04.020

Foster, N. E. V., & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex, 20*(6), 1350-1359. doi:10.1093/cercor/bhp199

Fraschini, M., & Hillebrand, A. (2017). Phase transfer entropy in matlab. Retrieved from https://figshare.com/articles/Phase_Transfer_Entropy/3847086

Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *The Journal of Neuroscience, 35*(45), 15187-15198. doi:10.1523/jneurosci.2397-15.2015

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *The Journal of Neuroscience, 32*(5), 1791-1802. doi:10.1523/jneurosci.4107-11.2012

Gelding, R. W., & Sun, Y. (2018). Commentary: Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced β-band oscillations during perception. *Frontiers in Neuroscience, 12*(50). doi:10.3389/fnins.2018.00050

Gelding, R. W., Thompson, W. F., & Johnson, B. W. (2015). The pitch imagery arrow task: effects of musical training, vividness, and mental control. *PLoS ONE, 10*(3), e0121809. doi:10.1371/journal.pone.0121809

Halpern, A. R. (2003). Cerebral substrates of musical imagery. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 217-230). New York, NY: Oxford University Press; US.

Halpern, A. R. (2015). Differences in auditory imagery self-reported predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain, 25*(1), 37-47. doi:10.1037/pmu0000081

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*(7), 697-704. doi:10.1093/cercor/9.7.697

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience, 24*(6), 1382-1397. doi:10.1162/jocn_a_00216

Hillebrand, A., Tewarie, P., van Dellen, E., Yu, M., Carbo, E. W. S., Douw, L., . . . Stam, C. J. (2016). Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proceedings of the National Academy of Sciences, 113*, 3867-3872.

Hubbard, T. L. (2013). Auditory aspects of auditory imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory Imagery* (pp. 51 - 76). S.I.: Springer New York.

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. doi:10.1016/j.cognition.2014.10.018

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences, 1169*(1), 58-73. doi:10.1111/j.1749-6632.2009.04579.x

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology, 245*, 15-26. doi:10.1016/j.expneurol.2012.09.014

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin, 129*(5), 723-746.

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *The Journal of Neuroscience, 29*(8), 2477-2485. doi:10.1523/jneurosci.4921-08.2009

Levitin, D. J., & Grafton, S. T. (2016). Measuring the representational space of music with fMRI: a case study with Sting. *Neurocase, 22*(6), 548-557. doi:10.1080/13554794.2016.1216572

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., . . . Scott, S. K. (2015). Feel the noise: relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex, 25*(11), 4638-4650. doi:10.1093/cercor/bhv134

Lobier, M., Siebenhuhner, F., Palva, S., & Palva, J. M. (2014). Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *NeuroImage, 85, Part 2*, 853-872. doi:10.1016/j.neuroimage.2013.08.056

McAuley, J. D., & Henry, M. J. (2010). Modality effects in rhythm processing: Auditory encoding of visual rhythms is neither obligatory nor automatic. *Attention, Perception, & Psychophysics, 72*(5), 1377-1389. doi:10.3758/app.72.5.1377

Meijer, D., te Woerd, E., & Praamstra, P. (2016). Timing of beta oscillatory synchronization and temporal prediction of upcoming stimuli. *NeuroImage, 138*, 233-241. doi:10.1016/j.neuroimage.2016.05.071

Moore, M. E. (2011). Imagination and the mind's ear. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 71*(11-A), 4046.

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences, 114*(42), E8913-E8921. doi:10.1073/pnas.1705373114

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex, 22*(11), 2466-2477. doi:10.1093/cercor/bhr325

Otto, F. (1990). *Glenn Gould: a life and variations*. New York: Vintage Books.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the action simulation for auditory prediction (ASAP) hypothesis. *Frontiers in systems neuroscience, 8*, 57. doi:10.3389/fnsys.2014.00057

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*(6), 718-724. doi:10.1038/nn.2331

Ross, B., Barat, M., & Fujioka, T. (2017). Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced beta-band oscillations during perception. *The Journal of Neuroscience, 37*(24), 5948-5959. doi:10.1523/jneurosci.3613-16.2017

Schaal, N. K., Pollok, B., & Banissy, M. J. (2017). Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Scientific Reports, 7*, 42456. doi:10.1038/srep42456

Schaal, N. K., Williamson, V. J., & Banissy, M. J. (2013). Anodal transcranial direct current stimulation over the supramarginal gyrus facilitates pitch memory. *European Journal of Neuroscience, 38*(10), 3513-3518. doi:doi:10.1111/ejn.12344

Schaal, N. K., Williamson, V. J., Kelly, M., Muggleton, N. G., Pollok, B., Krause, V., & Banissy, M. J. (2015). A causal involvement of the left supramarginal gyrus during the retention of musical pitches. *Cortex, 64*, 310-317. doi:10.1016/j.cortex.2014.11.011

Schaefer, R. S. (2017). Music in the brain: imagery and memory. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition*. Abingdon: Routledge.

Scott, D. W. (1992). *Multivariate density estimation : theory, practice, and visualization / David W. Scott*. New York: New York : Wiley.

Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research, 24*(1), 117-126. doi:10.1016/j.cogbrainres.2004.12.014

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017). Neural entrainment to the beat: the "missing-pulse" phenomenon. *The Journal of Neuroscience, 37*(26), 6331-6341. doi:10.1523/jneurosci.2500-16.2017

Tang, H., Crain, S., & Johnson, B. W. (2016). Dual temporal encoding mechanisms in human auditory cortex: Evidence from MEG and EEG. *NeuroImage, 128*, 32-43. doi:10.1016/j.neuroimage.2015.12.053

Zatorre, R. J. (2012). Beyond auditory cortex: working with musical thoughts. *Annals of the New York Academy of Sciences, 1252*(1), 222-228. doi:10.1111/j.1749-6632.2011.06437.x

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547-558.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron, 47*(1), 9-12. doi:10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22*(4), 775-789. doi:10.1162/jocn.2009.21239

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience, 37*(9), 1421-1434. doi:10.1111/ejn.12140

Chapter 6

Sensorimotor Contributions to Mental Imagery of Musical Rhythm

Rebecca W. **Gelding**[a,c], William F. **Thompson**[b,c], Blake W. **Johnson**[a,c]

[a] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[b] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

[c] ARC Centre of Excellence in Cognition and its Disorders

Author Contributions

All authors conceived and designed the experiment. R.W.G. conducted the experiments. R.W.G. & B.W.J. analysed the data and discussed the results. All authors wrote and edited the manuscript.

Abstract

Mental imagery of rhythm has been relatively unexplored, partly due to the difficulty of maintaining tempo in silence. To overcome this, a new Rhythm Imagery Task (RIT) was developed. This task required participants to mentally maintain simple piano rhythms, in time with a bass drum metronome, for a continuation period. Accuracy of imagery was objectively measured by requiring participants to respond to an audible probe that was played either in time or out of time with one of the imagined piano notes. A motor version of the task required participants to tap out the piano rhythm during the continuation period. Participants completed 40 Imagery trials, then 40 Tapping trials, followed by another 40 Imagery trials. Firstly, we validated the task behaviourally with 52 participants, finding that participants improved their imagery accuracy after short-term tapping training, even when tapping white noise masked the sound consequences of the taps. Next, using magnetoencephalography (MEG), 19 new participants were tested on the same paradigm. The β-band time course was extracted from bilateral auditory and sensorimotor sources, and modulation at the bass drum frequency (0.625 Hz) and the beat frequency (2.5 Hz) was compared. Greater β-band modulation was seen in the sensorimotor than auditory regions at the bass drum frequency for both heard and imagined patterns. The effect of tapping was seen most strongly in changes to directed connectivity in θ-band from the right sensorimotor region. More accurate imagery performance after tapping was related to increased right sensorimotor to left auditory driven connectivity during the silent on-beat in the syncopated patterns, both heard and imagined. Taken together, these results support the conclusion that precise timing provided by the right sensorimotor region facilitates accurate imagery of rhythm.

6.1 Introduction

Rhythm and pitch are considered the two primary dimensions of music (Krumhansl, 2000), one temporal and the other spectral. Musical rhythm is defined as the pattern of time intervals in a sequence (Grahn, 2012), which encapsulates tempo, meter and patterns of timing. Several studies of music cognition have indicated that the perceptual mechanisms of pitch and rhythm may be dissociable, with selective impairments possible in one domain not but not the other (Hyde & Peretz, 2004; Phillips-Silver et al., 2011). Pitch and timing of musical imagery are also dissociable as evidenced by behavioural tasks of musical imagery, where temporal accuracy is worse than pitch accuracy (Janata & Paroo, 2006; Weir et al., 2015).

While different spectral aspects of music including pitch, timbre, and loudness have been investigated in musical imagery studies, fewer studies have investigated temporal aspects of imagery. Most of these have investigated tempo and meter rather than rhythmic patterns (Hubbard, 2013). Imagery for tempo and changes in tempo have been explored using adjustable metronomes (Halpern, 1992) and tapping tasks (Jakubowski et al., 2018). In a study by Colley et al. (2018) participants were presented with 5 tones that were either speeding up or slowing down. After a gap of 2 beats, participants were required to respond to a probe tone that was presented too early or too late for a given temporal pattern. The gap between the timing of the probe and the correct answer was reduced over trials such that a discrimination threshold was reached when performance was at chance. Performance on this tempo imagery task was a significant predictor of performance on two behavioural tasks involving expressively timed music: A sensorimotor synchronisation task in which participants tapped along with the beat of the music; and a beat alignment task in which participants judged whether or not a click track was playing in time to the beat of the excerpts. Participants also completed a modified

version of the Pitch Imagery Arrow Task (Gelding et al., 2015) however pitch imagery performance, did not predict beat alignment performance. Hence Colley et al. (2018) concluded that tempo imagery is dissociable from pitch imagery, concordant with the results described above for pitch and rhythm perception.

Metrical hierarchy of beats has been explored in several studies requiring participants to imagine different metrical structures on top of heard beats and comparing patterns of brain activity for imagined accents and actual accents. Imagined accents can be decoded successfully using brain activity from perceived accents and vice versa, suggesting considerable overlap in processing of perceived and imagined metrical structure (Vlek et al., 2011). In addition, both heard and imagined accents on downbeats show greater β-band desynchronisation than for unaccented beats (Fujioka et al., 2015). The β-rhythm has also been shown to be more strongly modulated during the omission of accented vs unaccented tones (Snyder & Large, 2005). Further, when participants were required to imagine a beat to a syncopated pattern, the induced β-band was found to be modulated according to the imagined beat, not the auditory notes that occurred off the beat (Iversen et al., 2009). Right auditory cortex activity was shown to track the beat in syncopated rhythms in which there is no energy at the beat frequency, even when participants were not instructed to maintain the beat while listening (Tal et al., 2017).

Of the few studies that have explored imagery of rhythmic patterns, both manipulation and maintenance paradigms have been used. Zatorre et al. (2010) reported that mental reversal of familiar melodies was associated with activation of the intraparietal sulcus. However since this study involved mental manipulation of both rhythm and pitch attributes of melodies, it is unclear if there are mechanisms that may be specific to the manipulation of each of these dimensions. Stupacher et al. (2017) used a complex rhythm task in which participants had to maintain the rhythmic pattern in order

to judge the timing of a probe during a silent interval. In this task musicians showed enhancement of brain rhythms at the beat frequency, and the amplitude of beat-related rhythms was positively correlated with task accuracy. This design is promising, but the long silent periods of this task are likely to result in a gradual drift of the internal tempos.

The role of neuromotor systems in tracking tempo has received considerable attention (Manning et al., 2017; Manning & Schutz, 2013, 2015, 2016). These studies are relevant because their control tasks required ongoing rhythm imagery to complete successfully, although they were described simply as "no-movement" conditions. In these studies participants heard a simple isochronous pattern with an accent on every 4th beat (downbeat). After the 4th downbeat the subsequent beats were not sounded, but the next downbeat came in either slightly early, late or on time (Manning & Schutz, 2013). Tapping with a drumstick rather than finger tapping (i.e. using more motor effectors) led to better performance (Manning et al., 2017), and these improvements were maintained when auditory feedback was eliminated by masking with white noise (Manning & Schutz, 2015). When comparing results of the "no-movement" conditions in these tasks, there were no differences between percussionists and non-percussionists (Manning & Schutz, 2016), suggesting that tapping experience may not be a strong predictor of tempo imagery ability. However due to the interleaved block design of movement and "no-movement" conditions, it is not possible to investigate the precise role of tapping on imagery

Activity of and connectivity between auditory and sensorimotor regions has also been shown to be affected by both short-term motor training and long-term motor experience (Krishnan et al., 2018; Ross et al., 2017; Zatorre et al., 2007). Passive listening to a Tibetan singing bowl after a short period of learning to strike the bowl with the right hand, showed greater θ-phase connectivity between the right auditory and the left sensorimotor region (Ross et al., 2017). Such changes appear to be dependent on the

modality of training, as recent evidence has shown beat-boxers and guitarists show enhanced sensorimotor activation when passively listening to unfamiliar beat-box and guitar music respectively (Krishnan et al., 2018). This suggests that internal motor models are generated when participants listen to sounds they are experienced at producing (Krishnan et al., 2018). It remains unclear whether short-term tapping training on a rhythmic task would also cause neuroplastic changes during rhythm imagery.

In the present study we designed a novel Rhythm Imagery Task (RIT) to address these gaps in the literature. The RIT required participants to maintain simple rhythmic piano rhythms in their minds in time with a bass drum metronome. Imagery performance was objectively measured by requiring participants to determine whether an audible probe was played in time or out of time with one of the imagined piano notes. A "motor" version of the task required participants to physically tap out the piano rhythm during the continuation period. We first examined RIT performance comprehensively with a behavioural study, and then carried out a MEG study designed to elucidate the neural mechanisms associated with rhythm imagery.

Based on the results of Manning and Schutz (2015) we hypothesised that tapping would improve performance (i.e. performance accuracy in Block 2 of imagery would be greater than Block 1) and that participants who tapped in silence would improve more than those who tapped in white noise. From the results of Ross et al. (2017), we also predicted that tapping would increase directed connectivity between sensorimotor and auditory brain regions, as indexed by MEG measurements of brain activity during imagery and perception of rhythmic patterns.

6.2 Study 1: Behavioural Study

The two aims of this behavioural study were to test: (1) whether rhythm production improves rhythm imagery performance by comparing RIT performance before and after a block of tapping; and (2) the effect that auditory feedback of the tapping has on subsequent rhythm imagery performance. We predicted that both groups of participants would improve after short-term motor production, and that participants who tapped in silence would improve more than those who tapped in white noise.

6.2.1 Materials and Methods

6.2.1.1 Participants

52 participants were split into two experimental groups, tapping in silence ('Silence') or tapping in white noise to mask auditory feedback from the taps ('Noise'). 50 participants were right-handed and 2 participants were left-handed. All participants reported having normal or corrected-to-normal vision and normal hearing. All participants provided written consent and all procedures were approved by the Macquarie University Human Research Ethics Committee. After data collection, 4 participants were excluded because they tapped during the imagery condition or tapped out the rhythmic patterns incorrectly (see below for more details of the exclusion criteria). Of the remaining participants, 25 were in the Silence condition and 23 were in the Noise condition. All participants completed the Goldsmith Musical Sophistication Index (Gold-MSI; Mullensiefen et al., 2014) musical ability subscales, and the Bucknell Auditory Imagery Scale, with subscales for vividness (BAIS-V) and control (BAIS-C) (Halpern, 2015).

The characteristics of the two groups are summarised in Table 6-1. There were no significant group differences for any of variables except gender ($p < 0.04$), with more females in the silent group.

Table 6-1

Participant Demographics and Psychometric Scores (Behavioural Study)

			Gold-MSI Subscales						BAIS Subscales	
	Age	**No. of Females**	**Active Engagement**	**Perceptual Abilities**	**Musical Training**	**Emotions**	**Singing Abilities**	**General Sophistication**	**Vividness**	**Control**
Silence Tapping (N = 25)	24.5 (18 – 53)	21	39.4 (20 – 55)	48.6 (38 – 63)	30.2 (7 – 43)	33.8 (25 – 40)	33.28 (23 – 46)	84 (60 – 114)	4.9 (1.5 – 7)	5.2 (1.8 – 6.9)
Noise Tapping (N = 23)	21.1 (17 – 46)	13	40.7 (21 – 61)	50.8 (37 – 63)	30.7 (10 – 48)	32.9 (20 – 42)	35.2 (26 – 49)	87 (55 – 120)	5.1 (3.8 – 6.1)	5.2 (4 – 6)

6.2.1.2 Stimuli and Apparatus

Rhythm Imagery Task

A rhythmic pattern for each trial was made from the combination of two simple rhythmic units: a syncopated and a non-syncopated rhythm (see Figure 6-1). Each unit consisted of a bass drum on the first beat of the bar (downbeat), and piano tones, on 2 of the other beats in a 4-beat bar (as seen in Figure 6-1), with a constant beat rate of 400 ms. Trials began with a white fixation cross for 1.6 s, which remained on the screen until the end of the trial. A rhythmic pattern was played twice at the start of the trial, and then the piano component dropped out and the bass drum continued at a constant rate (1 per 1.6 seconds). The tempo remained constant throughout the whole task. Participants were instructed to imagine the sound of the piano pattern during this continuation phase,

keeping timing as accurately as possible. They were also instructed not to move or sing in any way to help them imagine, but to remain as still as possible.

In standard trials the whole pattern was imagined at least 3 times before a single piano tone probe was played. This piano tone was either in or out of time with one of the expected piano notes in the rhythmic pattern. Across the whole task there were 4 differently timed probes (see orange box, Figure 6-1). 50% of probes were in-time (correct) and 50% were out of time (incorrect). Half of incorrect probes (25% of trials) were on the rest beat: beat 2 of the unsyncopated unit (at 400 ms) or beat 3 of the syncopated unit ("on-beat", at 800 ms). The remaining incorrect probes (25% of trials) were midway between beat 3 and 4 (at 1000 ms) for both units. The probe was presented with equal probability during the first or second unit of the pattern. The probe was the same the piano note used in the trial.

Once the probe was played, the white fixation cross turned blue, and participants had to indicate with a button press whether the probe was in time or out of time with the imagined pattern. Feedback was given on every trial and the blue fixation cross turned green for correct responses and red for incorrect responses. To ensure that participants were maintaining the rhythmic pattern in imagery throughout the silent period, on 10% of trials (catch trials) the probe was played after maintaining the pattern just once. Pilot testing showed that active imagery was required for accurate task performance over a maintenance period of 3 repetitions (9.6 seconds).

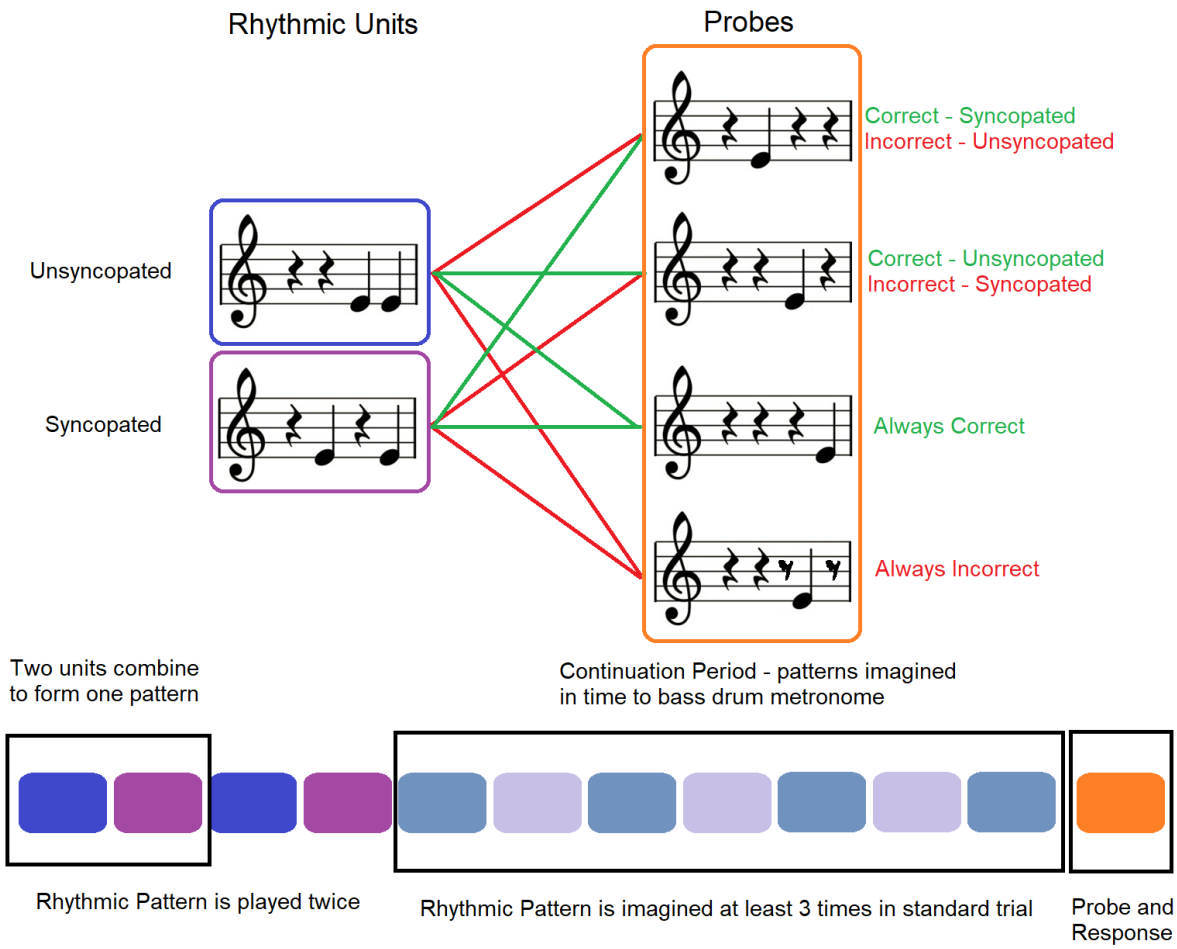

Figure 6-1: Schematic of a RIT trial. Each rhythmic pattern was a combination of two possible units (Unsyncopated, blue and/or Syncopated, purple), both of which had a bass drum on the first beat of each unit. Each rhythmic pattern was heard twice, then during the continuation period participants imagined the patterns along with a sounded bass drum on the first beat. In a standard trial, after 6 or 7 units, when the rhythmic pattern had been imagined at least 3 times, one of 4 possible probes (orange) was sounded, and participants were required to indicate if the probe was correct (green options) or incorrect (red options).

Given the 2 rhythmic units (Syncopated (S) and Unsyncopated (U)), there were 4 possible rhythmic pattern combinations (SS, SU, US, UU). For each of these, one of 4 possible probes (orange box in Figure 6-1) was presented either during the first or second

unit of the pattern (that is after 6 or 7 units have been imagined). For example, in Figure 6-1, the "US" pattern is imagined for 7 units, which means the participants should be imagining the "S" unit when the probe (orange box) is played. For standard trials there was equal probability of the probe falling on the first or second unit, hence for each pattern there are 4 possible probes for the first unit and 4 possible probes for the second, making 8 possible probe combinations for each of the 4 patterns (8 x 4 = 32 trials), of which half were correct and half were incorrect. Each block also contained 8 catch trials, where the probe was played earlier in the continuation period, after maintaining the pattern just once. The piano tone used in trials varied between C4, D4, E4, F4 and G4, but was constant within a given trial. Rhythmic patterns and pitch for a given trial were selected quasi-randomly so that there was no direct repeat of any pitch or rhythmic pattern presented.

Rhythm Tapping Task

In this task a white fixation cross appeared on the screen, and then participants were introduced to one of the simple rhythmic patterns of the RIT, that featured a bass drum (on downbeat) and piano tones. After hearing the pattern just once, the fixation cross then turned orange, and participants then tapped along *with* the piano part on a designated keyboard button (down arrow, which was labelled with an orange cross), as the pattern repeated. Then the piano part dropped out and the bass drum continued to keep participants in time. They were instructed to continue tapping out the piano pattern on the down arrow keyboard button.

In standard trials, the total pattern was tapped out at least three times (minimum of 9.6 seconds), and during the fourth repetition of the tapping at some point in the pattern a single piano note was played as a probe, that was either in time with one of the piano notes in the pattern or was out of time. After a 500 ms pause, a blue fixation cross

appeared, indicating that participants had to choose if the probe note was in time with the pattern they were playing or out of time. Feedback was given on every trial and the cross turned green for correct responses and red for incorrect responses. Forty trials were presented, with an opportunity for a break after 20 trials. For catch trials (20%), the probe was introduced after just one complete tapping of the pattern. These were included to ensure that participants maintained vigilance throughout the task for the probe.

For participants in the "Noise" group, white noise was inserted during the continuation period to mask the auditory consequences of tapping. The bass drum was audible over the noise. The masking noise stopped exactly when the probe tone was played.

6.2.1.3 Procedure

Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) was used to control the experiment and record responses. Audio files were created in Audacity (www.audacityteam.org) using the *Alicia's Keys* piano plugin (Native Instruments GmbH) and the bass drum from the EZDRUMMER™ 2 (www.toontrack.com) software and exported as .wav files for use in Presentation®.

Upon being seated in front of the computer, participants were given a sound check, whereby they could manually adjust the volume of their headphones to a suitable level. They were then introduced to the RIT and given two example trials. There were 40 RIT trials, with an opportunity for a rest after 20 trials. Participants then completed the Gold-MSI survey. They were then introduced to the tapping task and given two example trials. An opportunity for questions was given prior to the start of the task. The BAIS survey was then completed before starting the second block of RIT. After each condition, participants were asked to rate vividness of musical imagery of the piano tones on a scale

of 1 (not vivid and all) and 5 (very vivid), and to describe the strategies they used to complete the task.

6.2.1.4 Data Analysis

A 2 (Group: Silent, Noise) x 3 (Condition: RIT Block 1, Tapping, RIT Block 2) ANOVA was conducted for both accuracy and hit reaction time. A t-test compared Block 1 and Block 2 RIT for both accuracy and reaction time for all participants to determine if the tapping condition led to improved performance. Measures of tapping accuracy were calculated including absolute mean asynchrony of taps, as well as the coefficient of variation (CV), which is the standard deviation of the inter-onset interval (IOI) / mean IOI, for each of the inter-onset gaps (400 ms, 800 ms). Hence CV_400 is a measure of how consistently participants tapped the 400 ms gap found in the unsyncopated pattern and CV_800 is a measure of how consistently participants tapped the 800 ms gap in the syncopated pattern. Individual differences of musical sophistication (Gold-MSI), imagery vividness (BAIS-V) and control (BAIS-C) as well as measures of tapping accuracy were correlated with the performance measures (accuracy and reaction time) of the RIT Block 2 and tapping conditions. Multiple regression analysis used a hierarchical linear model through step-wise reduction with RIT accuracy from Block 2 as the criterion variable to determine significant predictors of rhythm imagery performance.

6.2.2 Results

6.2.2.1 Accuracy and Reaction Time

A 2 x 3 ANOVA of Accuracy revealed a significant main effect of Condition ($F_{(2,138)} = 7.392$, $p < .001$, $\eta^2 = .09$) and Group ($F_{(1,138)} = 4.096$, $p = .045$, $\eta^2 = .03$), but no significant interaction ($F_{(2,138)} = 0.593$, $p = .55$). Post hoc paired t-tests showed accuracy in tapping condition (M = 78.9%, SD = 15.1%) was better than both RIT Block

1 (M = 67.6%, SD = 16%; $t(47) = -7.782$, $p < .001$, 95% CI [-0.148, -0.087], Hedges g = .751) and RIT Block 2 (M = 73.8%, SD = 14.2%; $t(47) = 3.995$, $p < .001$, 95% CI [0.028, 0.084], Hedges g = .377). Crucially, performance was significantly more accurate in RIT Block 2 compared to RIT Block 1, confirming the hypothesis that short-term tapping training improves performance on the task ($t(47) = -3.742$, $p < .001$, 95% CI [-0.095, -0.029], Hedges g = .406).

Participants in the Noise condition (M = 76.2%, SD = 16.8%) out-performed participants in the Silent condition (M =71.2%, SD = 14.5%) as indicated by a significant main effect of Group. As Figure 6-2 illustrates, in RIT Block 1, even before the sound manipulation takes place in the Tapping Block, the Noise group had higher accuracy than the Silent group, however this difference was not statistically significant ($t(46) = -1.951$, $p = .057$, 95% CI [-0.178, 0.003], Hedges g = .554). As there was no interaction between Group and Condition, this suggests both groups show similar patterns of improvement after the tapping task, regardless of the auditory feedback manipulation. Results in the Noise Group show a ceiling effect for some participants.

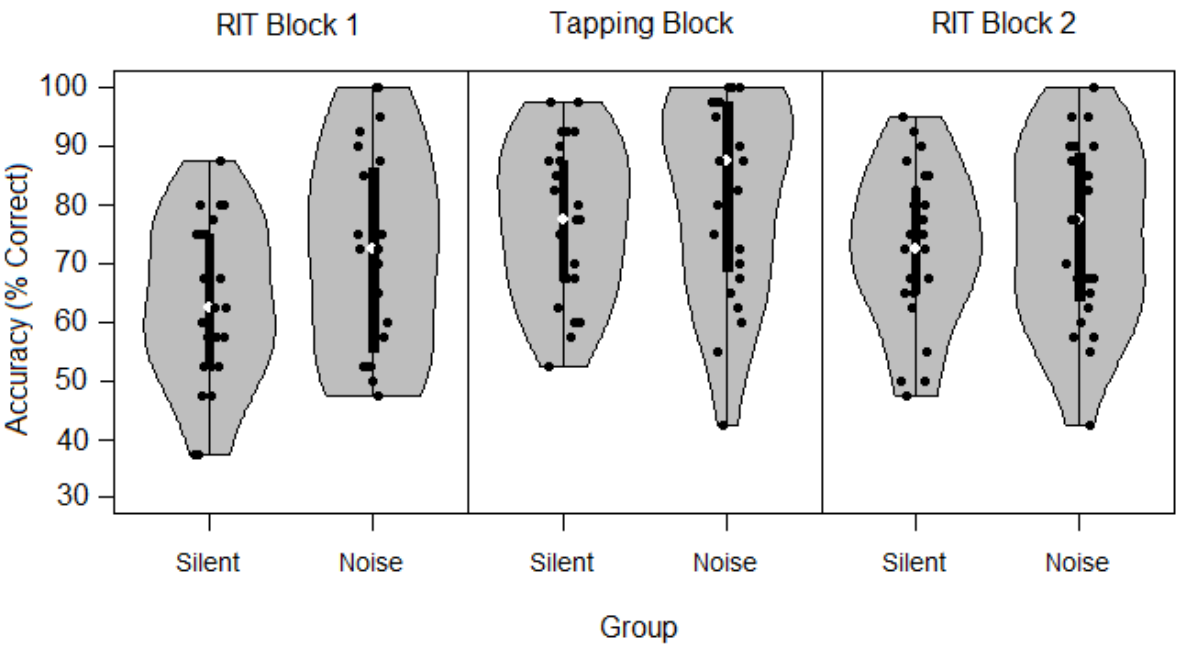

Figure 6-2: Performance Accuracy on each of the Blocks (RIT Block 1, Tapping, RIT Block 2) for the Silent and Noise groups separately. Median values of groups are white dots, black dots are individual accuracy scores. Solid black lines indicate 1st and 3rd quartiles, with the overall length of the violin plot showing maximum and minimum values per group.

A 2 x 3 ANOVA of mean hit reaction time revealed a significant main effect of Condition ($F_{(2,138)} = 13.07$, $p < .001$, $\eta^2 = .158$). Post hoc paired t-tests showed reaction times were faster in tapping (M = 466.6 ms, SD = 164.99 ms) than RIT Block 1 (M = 661.7 ms, SD = 220.27 ms; $t(47) = 6.961$, $p < .001$, 95% CI [138.74, 251.52], Hedges g = .995) and RIT Block 2 (M = 520.2 ms, SD = 186.54 ms; $t(47) = -2.56$, $p = .014$, 95% CI [-95.64, -11.49], Hedges g = .302). RIT Block 2 was also significantly faster than RIT Block 1 ($t(47) = 4.864$, $p < .001$, 95% CI [83.01, 200.12], Hedges g = .688).

6.2.2.2 Individual Differences

The scores on the Gold-MSI and BAIS subscales and three measures of tapping accuracy (Absolute Mean Asynchrony, CV_400 and CV_800) were correlated with accuracy (percent correct) and hit reaction time for tapping and RIT for Block 2 (see Table 6-2). The only significant correlations after Bonferroni correction were positive correlations for accuracy in both conditions, with both Musical Training and General Sophistication. In addition, there was a significant negative correlation between CV_800 and tapping accuracy. That is, participants with less variable and more consistent 800 ms gaps in the syncopated pattern (i.e., lower CV_800) showed higher tapping accuracy (percent correct).

6.2.2.3 Multiple Regression Analysis

Regression analyses were conducted to evaluate which variables best predicted accurate performance on the RIT in Block 2, taking percent correct as the criterion variable. Eighteen predictor variables (Percent correct RIT Block 1 and Tapping; Reaction times for RIT Block 1, Tapping, RIT Block 2; Absolute Mean Asynchrony; CV_400 CV_800; 6 Gold-MSI subscales, 2 BAIS subscales, Strategy Use from RIT Block 2 [1 = "Musical Imagery", 0 = "Alternative Strategy"]), Group [Noise = 1, Silent = 0]) were included in an initial model and stepwise regression reduced the model to the minimal adequate hierarchical linear model, with only significant predictors. This resulted in a final model containing only tapping accuracy and hit reaction time for Block 2 RIT, which together significantly predicted accuracy on RIT Block 2 ($F_{(2, 45)} = 43.18$, $p < .001$) and accounted for 66% of the variance in RIT accuracy ($R^2 = .657$; $R^2_{adj} = .642$). (See Table 6-3).

Table 6-2

Correlation of Key Variables with Accuracy and Reaction Time (RT) – Behavioural Study

	Gold-MSI Subscales						BAIS Subscales		Tapping Measures		
	Active Engagement	Perceptual Abilities	Musical Training	Emotions	Singing Ability	General Sophistication	BAIS-V	BAIS-C	Abs Mean Async	CV 400	CV 800
Tapping Accuracy	.281	.312	.603***	.206	.318	.511**	-.178	.046	-.413	-.414	-.602***
Tapping RT	.038	.174	.393	.187	.016	.186	.052	.184	.019	.091	-.137
RIT Block 2 Accuracy	.248	.295	.464*	.127	.348	.473*	-.211	.003	-.340	-.359	-.408
RIT Block 2 RT	.060	.131	.335	.159	.048	.163	.056	.026	-.039	.004	-.181

Correlations were Bonferroni corrected to account for multiple comparisons. Significance is denoted as ***= p< .001, **= p<.001, *= p <.05.

Table 6-3

Multiple Regression Results

Variable	β	SE	*t*	*p*
(Intercept)	0.234	0.075	3.106	0.003
Tapping Accuracy	0.739	0.082	8.958	<0.0001***
RIT Block 2 RT	-0.0002	0.00007	-2.38	0.022*
R^2_{adj}; R^2	.642; .657			
	F[2,45] = 42.18***			

Significance is denoted as * = p < .05, *** = p < .001.

6.2.3 Discussion

The two main predictions of this behavioural study were that tapping would improve performance, and that participants who tapped in silence would improve more than those who tapped in white noise. Our results confirm the first prediction, with both higher performance accuracy and faster reaction times in the RIT of Block 2 compared to Block 1 across all participants. In addition, the multiple regression analysis revealed that performance accuracy in Block 2 of the RIT was best predicted by tapping accuracy (percent correct) and reaction time in Block 2 RIT. Taken together this suggests that tapping significantly improved task performance.

However, the second prediction was not confirmed as the Silent condition participants did not out-perform Noise condition participants. In fact, overall, there was a significant main effect of Group in the ANOVA for performance accuracy, with the Noise condition group performing more accurately than the Silent condition participants. However, follow up t-tests revealed no significant difference between the individual blocks (RIT Block 1, Tapping or RIT Block 2). While participants in both groups were matched in musical training and auditory imagery vividness, the ceiling effect in the Noise group at RIT Block 1 meant there was less room for improvement after tapping in RIT Block 2. Hence the results do not conclusively address whether noise or silence during tapping leads to differential rhythm imagery performance. Even so, the present results do show both groups improve, suggesting that even in the absence of auditory feedback, tapping training helps enhance rhythm imagery performance.

These findings show that overt motor production during practice facilitated subsequent imagery performance and support the interpretation that the brain's motor

systems are involved in rhythm imagery (Manning et al., 2017; Manning & Schutz, 2015). In the following experiment, this interpretation was further evaluated with MEG measurements of auditory-sensorimotor brain function during rhythm imagery.

6.3 Study 2: MEG Study

The MEG study was designed to assess the relative activations and connectivity profiles of auditory and sensorimotor brain regions during rhythm imagery. We predicted that any plastic changes induced by practice with tapping should result in increased directed connectivity from sensorimotor to auditory brain regions.

6.3.1 Materials and Methods

6.3.1.1 Participants

19 participants were recruited. All had participated in a RIT screening task, which included additional rhythmic units than just the two featured in the present study and had completed the Gold-MSI and BAIS questionnaires. All participants obtained greater than 68% accuracy on the screening task and self-reported to being right-handed, having normal or corrected-to-normal vision and normal hearing. All participants provided written consent and procedures were approved by the Macquarie University Human Research Ethics Committee. Data from one participant who tapped overtly with their left hand during the imagery blocks was excluded.

6.3.1.2 Stimuli and Apparatus

The conditions of RIT Block 1, Tapping block and RIT Block 2 were identical to the behavioural study outlined above (Silent condition), except that no trial feedback was provided. Each block contained 40 trials, with a short break inserted after 20 trials. Between blocks participants had the opportunity to rest while marker coil measurements were taken.

6.3.1.3 MEG measurements

Data Acquisition

Brain activity was recorded with a whole-head MEG system (Model PQ1160R-N2, KIT, Kanazawa, Japan) consisting of 160 coaxial first-order gradiometers with a 50 mm baseline. Prior to MEG measurements, five marker coils were placed on the participant's head and their positions and the participant's head shape were measured with a pen digitiser (Polhemus Fastrack, Colchester, VT). Head position was measured by energising the marker coils in the MEG immediately before and after each block within the recording session. During acquisition MEG was sampled at 1 kHz and band-pass filtered between 0.03 and 200 Hz. Individual structural magnetic resonance images were not available for the present experiment so the adult template brain in BESA Research 6.1 (BESA Research, Gräfelfing, Germany) was used for all participants, using a spherical head model.

Responses were made on a MEG-compatible 4 button stick cylinder (Current Designs Inc: Philadelphia, USA, Model HHSC-1X4-CYL-2) with the right index finger. Three practice trials with feedback were administered inside the shielded room, before Block 1 (RIT) and the Tapping Block, but no practice trials were given in RIT Block 2. Functional localiser tasks (Arnal et al., 2015; Tal et al., 2017) after RIT Block 2. The auditory localiser was a random presentation of the same 5 piano tones used in the rhythm tasks. Tones were presented (n = 100) at a rate of 500 ms +/- 200 ms. The motor localiser tasks were self-paced button presses at a rate of approximately 1 per second for a total of 100 presses for each of the left and right index fingers. Participants were monitored with a video camera throughout. Upon completion participants were asked what strategies they used to complete the imagery task.

Surface electromyography (EMG) was monitored to rule out movements during the imagery task. EMG was measured with two pairs of bipolar electrodes attached to the forearm and underarm muscle, and recorded using BrainAmp ExG MR 16P (BrainProducts Gmbh, Gilching, Germany).

Behavioural Analysis

Accuracy was computed as the percent correct across completed trials. Hit reaction time (RT) was measured as the time between onset of the blue cross after the probe, indicating a response should be made, and the response. A t-test was calculated to confirm if rhythm imagery accuracy and RT improved between Block 1 and 2 of the RIT. Measures of tapping accuracy were also computed (Absolute mean asynchrony, CV_400, and CV_800). Individual differences in musical sophistication (Gold-MSI) and imagery vividness (BAIS-V) and control (BAIS-C) were compared to the performance measures (accuracy and reaction time) of RIT Block 2 and Tapping.

6.3.1.4 Source Localisation

MEG analyses were carried out in BESA Research 6.1 (BESA Research, Gräfelfing, Germany). Our spatial filtering approach used four dipole sources modelled in bilateral auditory cortices and bilateral sensorimotor cortices. Auditory sources were fit to each participant's data using the rising half of the averaged M100 response to the auditory localiser task. Average Talairach coordinates of the auditory sources in the BESA Research 6.1 standard template brain were x = 47.1 (right), y = -19 (posterior), and z = 3(superior) in the right hemisphere (Right Superior Temporal Gyrus, BA22) and x = -45.8, y = -20.4, and z = 3.6 in the left hemisphere (Left Superior Temporal Gyrus, BA22).

Bilateral sensorimotor sources were obtained for each participant, using a beamformer analysis of the β-band activity (15 – 30 Hz) for 300 ms around the onset of

the button presses from the motor localiser task, on left and right hand separately (n = 100 trials each). Bilateral sources within the sensorimotor cortices were found for each participant. The average Talairach coordinates of the BESA Research 6.1 standard template brain for these were: x = 38.9 (right), y = -19.8 (posterior), and z =47.7 (superior) in the right hemisphere (Right Postcentral Gyrus, BA3) and x = -36.9, y = -20.3, and z = 44.2 in the left hemisphere (Left Postcentral Gyrus, BA3).

The auditory and sensorimotor sources were then combined into a single model for each participant. The source localisation procedure resulted in the reduction of the 160 channel surface MEG data to a 4-source montage and provided spatial filters in close spatial proximity to bilateral auditory and sensorimotor cortices. Subsequent analyses were computed using the 4-source montage and using data only from trials with correct responses. Trials with MEG artefacts including blinks and eye-movements during baseline period -1100 – 0 ms were rejected from the time-frequency calculations using the artefact scan tool in BESA Research 6.1, which rejects trials based on abnormally high amplitudes (> 5000 μV) or abrupt rises or falls in amplitude (gradients > 2500).

Time-frequency plots were generated in BESA Research 6.1 using Temporal-Spectral Evolution using a frequency range of 4 to 50 Hz, a frequency sampling of 2 Hz and a time sampling of 25 ms over an epoch of -1100 ms – 16000 ms for standard trials. The baseline period (-1100 ms – 0 ms) had a silent white fixation cross on the screen. Two heard patterns (0 – 6400 ms) were then followed by 3 repeats of the imagined pattern (6400 – 16000 ms), while the bass drum played. The plots show the amplitude for each time point normalised to the mean amplitude of the baseline epoch for that frequency. A value of the time-frequency plot describes the spectral change of activity at time *t*, relative to the activity during the baseline epoch. A value of +100% means the amplitude is twice as high as during the baseline epoch. The evoked (averaged) signal was subtracted from

all trials prior to computing the mean time-frequency transform. Average time courses for induced β-band (16 – 26 Hz) were calculated in MATLAB 8.2 (MathWorks Inc, MA, USA). Analyses focussed on the β-band based on the results obtained in the pitch imagery MEG study (Chapter 5). While Chapter 5 used a 50 ms time bin, the present analyses used a smaller time bin of 25 ms in order to more precisely capture temporal evolution of rhythmic patterns. This required an increase of frequency bin size from 1 Hz to 2 Hz. Hence in the present study, the β-band was sampled over the range of 16 – 26 Hz, maintaining a band width of 10 Hz.

6.3.1.5 β-Band Modulation

The β-band activity was analysed in the frequency domain for modulations corresponding to the power spectrum of the acoustic stimuli, which had greatest power at the beat frequency (2.5 Hz) and the bass drum frequency (0.625 Hz). Fast Fourier Transforms were computed for the mean β-band time course for all rhythmic patterns (combination of 2 rhythmic units = 3.2 seconds total), averaged across both blocks. Repeated measures ANOVAs were computed for power obtained at 0.625 and 2.5 Hz.

6.3.1.6 Connectivity Analysis

Directed phase transfer entropy (dPTE) is a method that has recently been used to measure the effective connectivity between regions of interest (Hillebrand et al., 2016; Lobier et al., 2014). This method calculates the instantaneous phase of the time series from each region of interest, and, like Granger causality, determines the direction of information based on temporal precedence and influence of one region on another (Lobier et al., 2014). The advantage of this type of connectivity analysis is that it is model free, computationally straightforward and robust for various time windows and trial numbers (Lobier et al., 2014). To calculate dPTE values, for each participant, the raw data time courses for each unit (syncopated, unsyncopated), were epoched into the individual beats

of the bar (2^{nd}: 400 – 800 ms, 3^{rd}: 800 – 1200 ms, 4^{th}: 1200 – 1600 ms) and separately concatenated for imagined and heard patterns, for Block 1 and 2 of the RIT. The bass drum beat was excluded from analysis since it was heard in all patterns. These were then bandpass filtered in the θ-band (4 – 8 Hz). We focussed our analyses in the θ-band based on the results of Chapter 5, and because increased auditory-sensorimotor phase connectivity in this frequency band after motor training has been reported previously by Ross et al. (2017). dTPE was calculated using *PhaseTE_MF.m* (Fraschini & Hillebrand, 2017) using the bin size (h) of Scott (1992): $h = 3.49\sigma n^{-1/3}$ where σ is the standard deviation of the phase data calculated from the Hilbert transform for the time series, and n is the length (number of samples) of the time series. The default time delay was determined by the frequency content of the data. The results were then normalised between -0.5 and 0.5 with the sign indicating the direction of the connectivity, as per Morillon and Baillet (2017). This resulted in dPTE values per participant for each beat (2^{nd}, 3^{rd}, 4^{th}), each condition (heard, imagined) and each block (first, second) and each of the 6 connections between regions. dPTE values were then tested statistically with ANOVAs.

6.3.2 Results

6.3.2.1 Behavioural Results

Accuracy and Reaction Time

RIT accuracy was not significantly different between Blocks 1 and 2 (Block 1: M = 79.7%, SD = 13.5%; Block 2: M = 80.3%, SD = 12.5%; $t(17) = -0.229$, $p = 0.822$, 95% CI [-0.057, 0.046], Hedges $g = .042$). However hit reaction times were significantly faster after the tapping block (Block 1: M = 555.93 ms, SD = 172.09 ms, Block 2: M = 436.90 ms, SD = 119.05 ms, $t(17) = 3.437$, $p = .0031$, 95% CI [45.97, 192.10], Hedges $g = .042$).

Individual Differences

Scores on the subscales of the Gold-MSI and BAIS and three measures of tapping accuracy (absolute mean asynchrony, CV_400 and CV_800) were correlated with accuracy (percent correct) and hit reaction time for Tapping and RIT for Block 2. Correlations were Bonferroni corrected to account for multiple comparison, and no correlations survived correction. The largest correlations were between CV_800 and both Tapping accuracy and RIT Block 2 accuracy, however General Sophistication and Musical Training did not correlate strongly with performance on either task (see Table 6-4).

Table 6-4

Correlations of Key Variables from with Accuracy and Reaction Time (RT) – MEG Study

	Gold-MSI Subscales						**BAIS Subscales**		**Tapping Measures**		
	Active Engagement	**Perceptual Abilities**	**Musical Training**	**Emotions**	**Singing Ability**	**General Sophistication**	**BAIS-V**	**BAIS-C**	**Abs Mean Async**	**CV 400**	**CV 800**
Tapping Accuracy	.326	-.001	-.031	.152	.317	.195	.466	.321	-.398	-.230	-.564
Tapping RT	-.142	-.121	.006	-.195	-.459	-.259	-.439	-.476	.341	.028	.486
RIT Block 2 Accuracy	.142	-.149	.025	-.096	.117	.036	.329	-.009	-.321	-.393	-.531
RIT Block 2 RT	-.088	-.254	-.294	-.118	-.493	-.448	-.238	-.363	.254	-.124	.340

Correlations were Bonferroni corrected to account for multiple comparison and no correlations survived correction.

6.3.2.2 MEG Results

Normalised β-band Time Course

Normalised β-band time courses are plotted in Figure 6-3. The absolute area under the (rectified) curve from this figure showed greater mean β-modulation for heard units (Unsyncopated: 0.273%, Syncopated: 0.196%) than for imagined units (Unsyncopated: 0.171%, Syncopated: 0.143%).

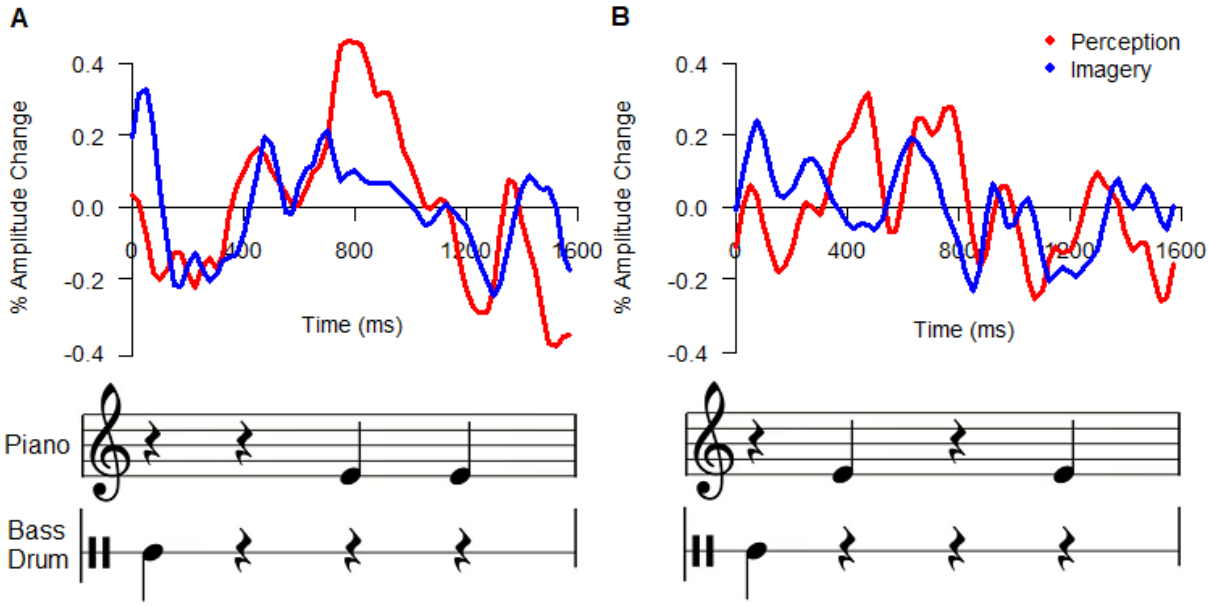

Figure 6-3: Mean β-band time course for 4 sources, for Perception (Red) and Imagery (Blue) of the **A)** Unsyncopated (piano tones: 800 ms, 1200 ms) and **B)** Syncopated (piano tones: 400 ms, 1200 ms) units.

Modulation at Bass Drum and Beat Frequencies

A repeated measures ANOVA of 2 Conditions (Imagery; Perception) x 2 Beat Frequencies (Bass Drum (0.625 Hz), Beat (2.5 Hz)) x 2 Sources (Auditory; Sensorimotor) x 2 Hemispheres (Right; Left) was computed across both blocks. A significant main effect was found for both Beat Frequency ($F_{(1,272)} = 38.53$, $p < .001$, $\eta_p^2 = .124$) and Condition ($F_{(1,272)} = 4.109$, $p = .0436$, $\eta_p^2 = .014$) with the bass drum frequency (0.625 Hz) having greater β-band power than the beat frequency (2.5 Hz), and heard patterns showing greater

modulation than imagined patterns. In addition there was a significant Hemisphere x Source interaction ($F_{(1,272)}$ = 4.645, p = .032, η_p^2= .017). Post hoc comparisons (Tukey HSD α = 0.05) revealed this was due to the right sensorimotor region having significantly greater power than right auditory in both conditions (p = .025).

To examine the main effect of Beat Frequency in more detail, additional ANOVAs (2 Conditions x 2 Sources x 2 Hemispheres) were computed for the bass drum frequency and beat frequency separately. For the bass drum frequency, the only significant main effect was Source ($F_{(1,136)}$ = 4.592, p = .034, η_p^2= .033), with sensorimotor regions showing greater modulation than auditory regions. At the beat frequency the only significant main effect was Condition ($F_{(1,136)}$ = 7.532, p = .007, η_p^2= .052), with heard patterns showing greater modulation than imagined.

To examine the main effect of Conditions in more detail, additional ANOVAs (2 Beat Frequencies x 2 Sources x 2 Hemispheres) were conducted for heard and imagined patterns separately. In the heard patterns, a significant main effect was found for beat frequency ($F_{(1,136)}$ = 12.881, p < .001, η_p^2= .087) confirming greater power at 0.625 Hz than 2.5 Hz. There was also a significant Hemisphere x Source interaction ($F_{(1,136)}$ = 5.570, p = .020, η_p^2= .039). However, post hoc comparisons (Tukey HSD α = 0.05) revealed none of these differences survive correction for multiple comparisons.

For imagined patterns, a main effect was seen for both beat frequency ($F_{(1,136)}$ = 30.330, p < .001, η_p^2= .182) and Source ($F_{(1,136)}$ = 4.669, p = .033, η_p^2= .033). There was also a significant Beat x Source interaction ($F_{(1,136)}$ = 4.669, p = .033, η_p^2= .033). Post hoc comparisons (Tukey HSD α = 0.05) revealed a significant difference at the 0.625 Hz frequency, with the sensorimotor sources showing greater β-band power than the auditory sources (p = .015), but no significant difference at the 2.5 Hz frequency between auditory

and sensorimotor sources (p = .999). The sensorimotor region bass frequency modulation was also greater than beat frequency modulation in both auditory (p < .001) and sensorimotor regions (p < .001).

In summary, the results showed greater β-band modulation at 0.625 Hz frequency than 2.5 Hz. At 0.625 Hz, sensorimotor sources showed greater β-band modulation than auditory sources during both heard and imagined patterns. At 2.5 Hz no Source differences were found, but heard patterns showed greater modulation than imagined patterns. As Figure 6-4 shows, the right auditory source for both the bass drum and beat frequency showed no significant difference between heard and imagined patterns.

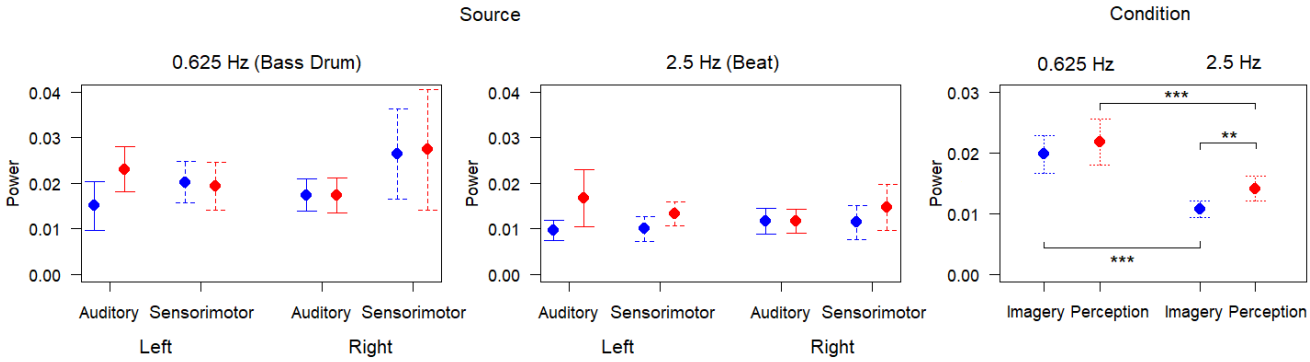

Figure 6-4: β-band power at the Bass Drum (0.625 Hz) and Beat (2.5 Hz) frequencies. 1st and 2nd plot show mean results and 95% Confidence Intervals for each of the four sources for the two conditions (Imagery: Blue; Perception: Red), with auditory sources having solid lines and sensorimotor sources having dashed lines. 3rd plot shows mean results and 95% Confidence Intervals of all sources for each of the two conditions for both frequencies. (** = $p < .01$, *** = $p < .001$)

Connectivity Analysis

The dPTE values were calculated for the two rhythmic units, for both heard and imagined patterns in both blocks separately (Block 1 before tapping and Block 2 after tapping), see Figure 6-5. To determine if there were differences in dPTE values before and after tapping, repeated measures ANOVAs were calculated for each of the 6

connections separately: 2 Condition (Heard, Imagined) x 2 Units (Syncopated, Unsyncopated) x 2 Block (1st, 2nd) x 3 Beats (2nd: 400 – 800 ms, 3rd: 800 – 1200 ms, 4th: 1200 – 1600 ms). Although tapping was completed by the right hand, and hence changes in connectivity would be expected between the left sensorimotor region and the auditory regions, both the right and the left auditory connection to the left sensorimotor region showed no significant effects of tapping. However the left auditory to the right sensorimotor connection showed a significant interaction effect of Block x Beat ($F_{(1,408)}$ = 3.050, p = .049, η_p^2= .015) as well as effects of Unit x Block ($F_{(1,408)}$ = 4.193, p = .041, η_p^2= .010), and Condition x Block x Beat ($F_{(2,408)}$ = 3.401, p = .034, η_p^2= .003). This was due to greater right sensorimotor to left auditory activity in imagined units during the 2nd beat in Block 1, and greater left auditory to right sensorimotor activity in heard units during the 4th beat in Block 2. This is to be expected given that in both syncopated and unsyncopated units, the 4th beat has a piano tone sounded in the heard units. In addition, the right auditory to the right sensorimotor connection also showed a significant interaction effect of Block x Beat ($F_{(1,408)}$ = 3.806, p = .023, η_p^2= .018), although post hoc comparisons (Tukey HSD α = 0.05) revealed no significant differences.

The hemispheric connection between the auditory regions showed a main effect for Unit x Beat ($F_{(2,408)}$ = 4.492, p = .012, η_p^2= .022), however post hoc Tukey HSD comparisons revealed no significant differences. Finally, the hemispheric connection between the sensorimotor regions showed significant interaction effects of Condition x Block ($F_{(1,408)}$ = 4.562, p = .033, η_p^2= .011), and Block x Beat ($F_{(2,408)}$ = 4.474, p = .012, η_p^2= .021). Post hoc Tukey HSD comparisons revealed in the first interaction, heard units showed greater left to right driven connectivity before tapping, but reversed direction from right to left driven connectivity afterwards (p = .044). The second interaction is due

to this same change from left to right driven connectivity before tapping to right to left driven connectivity seen in the on-beat (3rd beat: 800 – 1200 ms) (p = .011).

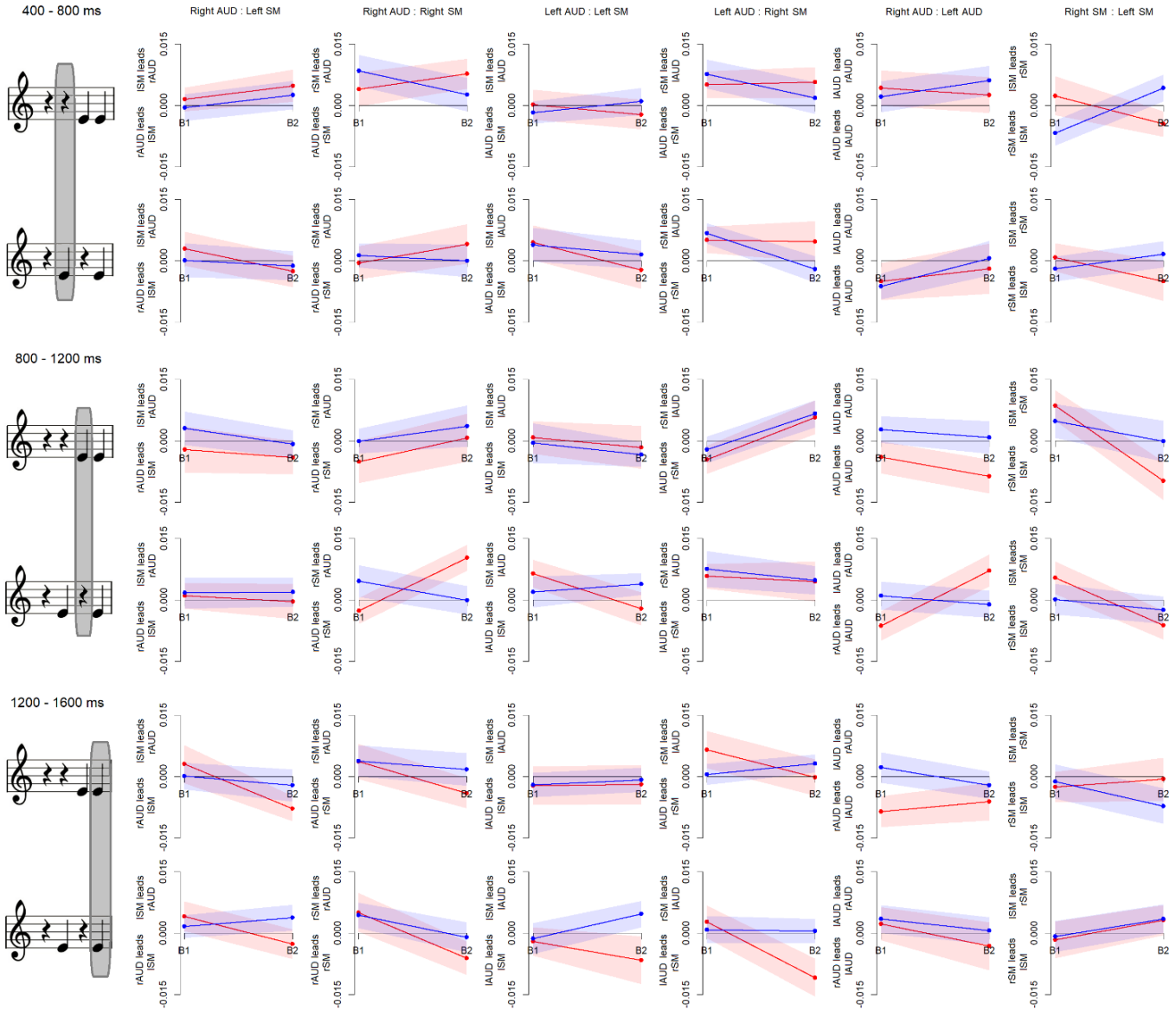

Figure 6-5: dPTE values for θ-band for RIT Block 1 (B1) and Block 2 (B2). Red is heard units; Blue is imagined units; shaded regions represent the standard error of the mean. Grey shaded region over music shows the time window of the row, corresponding to the musical beat. Columns 1 – 4 show sensorimotor regions as positive values and auditory regions as negative. Columns 5,6 show left regions as positive values and right regions as negative values.

Exploratory Analysis

Given the previous relationship found between imagery accuracy performance and directed connectivity from sensorimotor to auditory regions during temporal prediction (Chapter 5), correlations of the dPTE values with performance accuracy of RIT in Block 2 were also explored. The left auditory to right sensorimotor connection was chosen for this exploratory analysis as the right sensorimotor region showed greater modulation of the β-band at the bass drum frequency and showed significant differences between its connection to the left auditory region in the dPTE analysis. To compare with the results of Chapter 5, epochs were chosen in the syncopated and unsyncopated units during a rest beat. That is, the 2nd beat in the unsyncopated unit (top row, Figure 6-6) and the 3rd beat in the syncopated unit (bottom row, Figure 6-6). Correlations between RIT Block 2 accuracy and the dPTE values in both Block 1 and 2 for imagined and heard units were calculated, to determine if there was an effect of tapping. After FDR correction the only significant correlations found were for the Block 2 syncopated unit at the on-beat for both heard (r = .647, p = .023 FDR corrected) and imagined (r = .593, p = .038 FDR corrected) units. This suggests that participants with better rhythm imagery ability showed more right sensorimotor to left auditory driven activity after tapping during the silent on-beat. This association was not seen in the silent rest of the unsyncopated unit, or in the syncopated unit prior to tapping. This neuroplastic change at the on-beat was seen in both heard and imagined units.

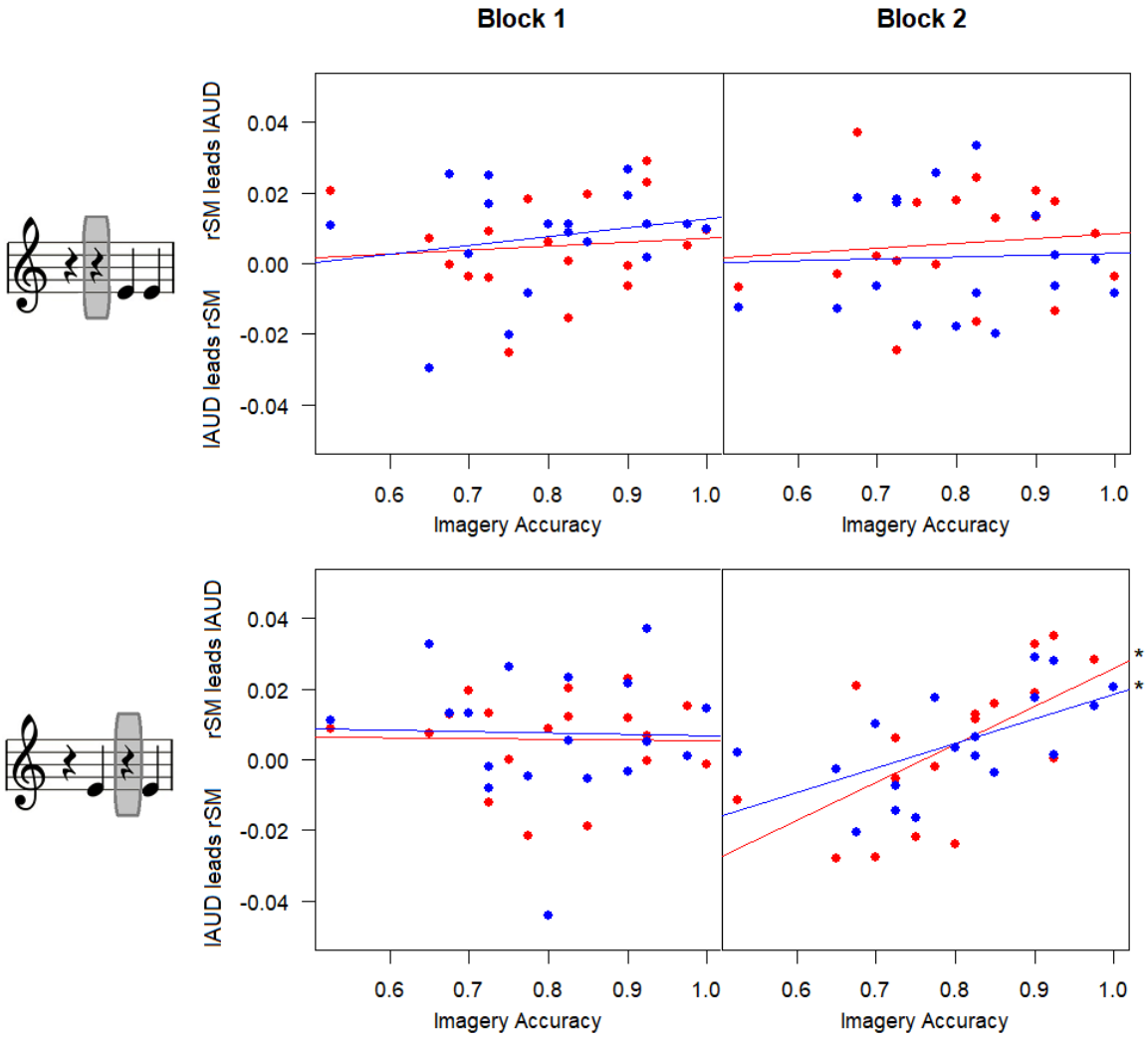

Figure 6-6: Correlations between Imagery accuracy in RIT Block 2 (percent correct) and dPTE values for the left auditory (lAUD) to right sensorimotor (rSM) connection in both heard (red) and imagined (blue) units in Block 1 (first column) and Block 2 (second column) during silent rests. Positive values indicate more sensorimotor driven activity, negative values indicate more auditory driven activity. Top row are unsyncopated units, bottom row are syncopated units. The only significant correlations after FDR correction were for both heard and imagined units in Block 2 for the syncopated rhythm (bottom right; * = $p < .05$ FDR corrected)

6.3.3 Discussion

The results of this MEG study support the interpretation that tapping modulates sensorimotor to auditory directed θ-band connectivity, for both imagined and heard

patterns. In particular, results showed that participants with more accurate imagery show an increase in the right sensorimotor to left auditory directed connection during the silent on-beat in a syncopated rhythmic pattern. During passive listening, increased θ-band functional connectivity has previously been seen between the left sensorimotor and right auditory regions after short-term motor training (Ross et al., 2017). However that used study weighted Phase Lagging Index which does not provide directionality of connectivity (Gelding & Sun, 2018). The advantage of using dPTE is that it provides evidence for the direction of phase activity between sources (Lobier et al., 2014).

In addition, the results of this study show 4 salient features of β-band modulation during imagery and perception of rhythms. First, for both heard and imagined patterns, β-band modulation in the sensorimotor regions was greater than auditory regions at the bass drum frequency (0.625 Hz), particularly the right sensorimotor region. As the bass drum frequency marks an important time point in the hierarchical structure of the pattern, namely the downbeat at the onset of the bar, this is consistent with the proposal that motor regions are involved in beat perception (Grahn & Brett, 2007). Second, for both heard and imagined patterns, the β-band modulation at the bass drum frequency was greater than the beat frequency. Recent studies have shown that rhythmic stimulation by bass sounds leads to enhanced neural representation of the beat and meter (Lenc, Keller, Varlet, & Nozaradan, 2018), hence this effect may be due to the use of the bass drum rather than piano tone at the start of each bar. The bass drum was chosen to be more ecologically valid as a marker of beat, and hence help participants in their task of maintaining tempo for their imagery. An alternative interpretation of this result is that the bass drum marks an important strong beat in the metrical hierarchy, and supports the notion that the β-band is modulated more to an imagined beat rather than auditory notes on the off-beat in a syncopated pattern (Iversen et al., 2009).

Third, at the beat frequency (2.5 Hz) β-band modulation was greater for heard patterns than imagined patterns. This finding stands in contrast to the pitch imagery results described in Chapter 5, showing greater β-band modulation during imagery than perception. It is possible that this result may be due to less precise timing of imagined beats, resulting in less power at the 2.5 Hz frequency. Alternatively, given recent evidence that the β-band desynchronisation may be related to upcoming predictability of the following tone (Chang et al., 2018), it may reflect the fact that in the heard rhythmic patterns, the timing of the onset of the next heard note was more predictable than in the imagined condition, where participants were anticipating hearing a probe tone at any time, whilst continuing to imagine the pattern in their minds.

Fourth, at the 2.5 Hz beat frequency no differences were found between the auditory and sensorimotor regions during imagined patterns. These results are consistent with Chapter 5 which also showed no differences between sources in the β-band modulation at 1 Hz (rate of imagined tone), during manipulation of pitch imagery. Hence this is further evidence that imagery involves the coordination of auditory and sensorimotor regions, consistent with the notion of the β-band being an “open line” of communication between the regions (Tang et al., 2016).

Taken together these results suggest the right sensorimotor cortex supports beat tracking during both heard and imagined patterns; and that the tapping manipulation produced neuroplastic changes in the form of enhanced connectivity from right sensorimotor to left auditory regions during the on-beat in syncopated patterns. In the β-band, sensorimotor regions modulate more than the auditory regions at higher levels of the metrical structure (that is, the bass drum), however at the beat/note level, no differences between the regions are seen suggesting a coordination of auditory and sensorimotor regions that enable imagery of rhythmic patterns.

6.4 General Discussion

The main results of these studies have shown that tapping enhances the involvement of the sensorimotor brain regions during subsequent imagery of rhythms. This is seen in the behavioural study, which showed that tapping practice improved subsequent rhythm imagery performance; and in the MEG study, where imagery accuracy was associated with increased directed connectivity between the right sensorimotor and left auditory cortices.

In the behavioural study, rhythm imagery accuracy significantly improved after a block of tapping, regardless of whether the auditory consequences of the taps were masked by white noise. This has been reported previously for maintenance of tempo (Manning & Schutz, 2015), but the present results are the first to show this effect for imagery of rhythmic patterns. In addition, imagery accuracy was best predicted by tapping accuracy and rhythm imagery reaction time, and both rhythm imagery and tapping accuracy positively correlated with musical training and general sophistication subscales of the Gold-MSI. Hence participants with more musical training were more accurate at imagining rhythmic patterns as well as tapping them out. Whilst Manning and Schutz (2016) found that percussionists were no better than non-musicians in a "no-movement" condition of tempo tracking, this result may have been due to the more difficult nature of their task, in which participants tracked the tempo for one bar, and judged the timing of probe tones that were on-time or late by either 75 ms or 150 ms. In comparison the RIT incorrect probes were offset by 200 ms or 400 ms.

Whilst the MEG study did not show an improvement in RIT accuracy after tapping, 2 significant changes related to behavioural performance did occur. Firstly, there was a significant improvement in hit reaction time after tapping for RIT trials. Secondly,

after tapping, a significant positive relationship emerged between imagery accuracy and the amount of directed connectivity in θ-band, between the right sensorimotor region and the left auditory region. This change occurred only for the on-beat (i.e. 3rd beat in the bar) in the syncopated pattern, where the pattern has a rest on a strong beat. Previous research had shown that passive listening to a Tibetan singing bowl after a short period of learning to strike the bowl with the right hand, showed greater θ-phase connectivity between the right auditory and left sensorimotor region (Ross et al., 2017). Our results suggest training in tapping out rhythmic patterns led to greater recruitment of the right sensorimotor region in tracking the strong beats in silence and points to similar mechanisms that are involved in beat perception for heard and imagined patterns.

The β-band modulation results suggest that the right sensorimotor region also plays an important role in tracking the bass drum frequency. Previously the right auditory cortex has been implicated in tracking the beat in syncopated rhythms in which there is no energy at the beat frequency (Tal et al., 2017), however in that task participants were passively listening to the patterns and were not instructed to maintain the beat. In the RIT maintaining the beat was explicitly required to competently perform the task.

Finally, the present results complement those seen in Chapter 5 which also showed a significant relationship with imagery accuracy and sensorimotor to auditory driven activity during musical imagery. However, in Chapter 5 pitch imagery accuracy was associated with increased left sensorimotor to right auditory regions, during the period just before the onset of a sound, that is, during temporal prediction. In the present MEG study rhythm imagery accuracy was associated with increased right sensorimotor to left auditory regions driven connectivity at the silent strong beat. This suggests left sensorimotor regions may be more involved in pitch imagery and right sensorimotor regions may be more involved in tracking the beat in rhythm imagery and perception.

Pitch and tempo imagery have previously been found to be dissociable (Jerde, Childs, Handy, Nagode, & Pardo, 2011; Schaal et al., 2017). Colley et al. (2018) found participants with better pitch imagery performance on a modified version of the PIAT (Gelding et al., 2015), had higher synchronisation scores when tapping along to expressive music with varying tempo but were not able to necessarily anticipate upcoming beat intervals. Rather, tempo imagery ability was found to be associated with prediction tracking of the beat (Colley et al., 2018). In addition, brain stimulation studies have shown pitch and rhythm memory are differentially affected when stimulating the supramarginal gyrus in each hemisphere, with the left hemisphere more responsible for pitch memory and the right hemisphere more responsible for rhythm memory (Schaal et al., 2017; Schaal, Williamson, et al., 2015).

Taken together these studies have introduced a novel rhythm imagery task and shown that performance on rhythm imagery improves after tapping, even in the absence of auditory feedback. Successful performance on the RIT requires precise timing of musical imagery. The MEG results show that the brain's sensorimotor systems, particularly the activity in the right sensorimotor region, are associated with accurate imagery of rhythms, and that the auditory-sensorimotor interactions show similar β-band modulation at the rhythmic beat level. Participants with better rhythm imagery ability showed increased right sensorimotor to left auditory driven connectivity during silent strong beats, which is consistent with the notion that beat perception involves implicit imagery.

6.5 References

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. Cerebral Cortex, 25(9), 3077-3085. doi:10.1093/cercor/bhu103

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2018). Beta oscillatory power modulation reflects the predictability of pitch change. *Cortex, 106*. doi:10.1016/j.cortex.2018.06.008

Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronization with expressively timed music. *The Quarterly Journal of Experimental Psychology, 71*(8), 1781 - 1796. doi:10.1080/17470218.2017.1366531

Fraschini, M., & Hillebrand, A. (2017). Phase transfer entropy in matlab. Retrieved from https://figshare.com/articles/Phase_Transfer_Entropy/3847086

Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *The Journal of Neuroscience, 35*(45), 15187-15198. doi:10.1523/jneurosci.2397-15.2015

Gelding, R. W., & Sun, Y. (2018). Commentary: Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced β-band oscillations during perception. *Frontiers in Neuroscience, 12*(50). doi:10.3389/fnins.2018.00050

Gelding, R. W., Thompson, W. F., & Johnson, B. W. (2015). The pitch imagery arrow task: effects of musical training, vividness, and mental control. *PLoS ONE, 10*(3), e0121809. doi:10.1371/journal.pone.0121809

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics In Cognitive Science, 4*(4), 585-606. doi:10.1111/j.1756-8765.2012.01213.x

Grahn, J. A., & Brett, M. (2007). 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

Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 1-27). Hillsdale, N.J.: L. Erlbaum Associates.

Halpern, A. R. (2015). Differences in auditory imagery self-reported predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain, 25*(1), 37-47. doi:10.1037/pmu0000081

Hillebrand, A., Tewarie, P., van Dellen, E., Yu, M., Carbo, E. W. S., Douw, L., . . . Stam, C. J. (2016). Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proceedings of the National Academy of Sciences, 113*, 3867-3872.

Hubbard, T. L. (2013). Auditory aspects of auditory imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory Imagery* (pp. 51 - 76). S.I.: Springer New York.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356-360. doi:10.1111/j.0956-7976.2004.00683.x

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences, 1169*(1), 58-73. doi:10.1111/j.1749-6632.2009.04579.x

Jakubowski, K., Bashir, Z., Farrugia, N., & Stewart, L. (2018). Involuntary and voluntary recall of musical memories: A comparison of temporal accuracy and emotional responses. *Memory & Cognition, 46*(5), 741-756. doi:10.3758/s13421-018-0792-x

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception & Psychophysics, 68*(5), 829-844. doi:10.3758/BF03193705

Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *NeuroImage, 57*(4), 1572-1579. doi:10.1016/j.neuroimage.2011.05.061

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., . . . Scott, S. K. (2018). Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play. *Cerebral Cortex, bhy208*. doi:10.1093/cercor/bhy208

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*(1), 159-179. doi:10.1037/0033-2909.126.1.159

Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1801421115

Lobier, M., Siebenhuhner, F., Palva, S., & Palva, J. M. (2014). Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *NeuroImage, 85, Part 2*, 853-872. doi:10.1016/j.neuroimage.2013.08.056

Manning, F. C., Harris, J., & Schutz, M. (2017). Temporal prediction abilities are mediated by motor effector and rhythmic expertise. *Experimental Brain Research, 235*(3), 861-871. doi:10.1007/s00221-016-4845-8

Manning, F. C., & Schutz, M. (2013). "Moving to the beat" improves timing perception. *Psychonomic bulletin & review, 20*(6), 1133-1139. doi:10.3758/s13423-013-0439-7

Manning, F. C., & Schutz, M. (2015). Movement enhances perceived timing in the absence of auditory feedback. *Timing & Time Perception, 3*(1-2), 3-12. doi:10.1163/22134468-03002037

Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research, 80*(4), 532-542. doi:10.1007/s00426-015-0678-5

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences, 114*(42), E8913-E8921. doi:10.1073/pnas.1705373114

Mullensiefen, D., Ginras, B., Musil, J. J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE, 9*(2), e89642. doi:10.1371/journal.pone.0089642

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia, 49*(5), 961-969. doi:10.1016/j.neuropsychologia.2011.02.002

Ross, B., Barat, M., & Fujioka, T. (2017). Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced beta-band oscillations during perception. *The Journal of Neuroscience, 37*(24), 5948-5959. doi:10.1523/jneurosci.3613-16.2017

Schaal, N. K., Pollok, B., & Banissy, M. J. (2017). Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Scientific Reports, 7*, 42456. doi:10.1038/srep42456

Schaal, N. K., Williamson, V. J., Kelly, M., Muggleton, N. G., Pollok, B., Krause, V., & Banissy, M. J. (2015). A causal involvement of the left supramarginal gyrus during the retention of musical pitches. *Cortex, 64*, 310-317. doi:10.1016/j.cortex.2014.11.011

Scott, D. W. (1992). *Multivariate density estimation : theory, practice, and visualization / David W. Scott*. New York: New York : Wiley.

Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research, 24*(1), 117-126. doi:10.1016/j.cogbrainres.2004.12.014

Stupacher, J., Wood, G., & Witte, M. (2017). Neural entrainment to polyrhythms: A comparison of musicians and non-musicians. *Frontiers in Neuroscience, 11*(208). doi:10.3389/fnins.2017.00208

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017). Neural entrainment to the beat: the "missing-pulse" phenomenon. *The Journal of Neuroscience, 37*(26), 6331-6341. doi:10.1523/jneurosci.2500-16.2017

Tang, H., Crain, S., & Johnson, B. W. (2016). Dual temporal encoding mechanisms in human auditory cortex: Evidence from MEG and EEG. *NeuroImage, 128*, 32-43. doi:10.1016/j.neuroimage.2015.12.053

Vlek, R. J., Schaefer, R. S., Gielen, C. C. A. M., Farquhar, J., & Desain, P. (2011). Shared mechanisms in perception and imagery of auditory accents. *Clinical Neurophysiology, 122*(8), 1526-1532.

Weir, G., Williamson, V. J., & Mullensiefen, D. (2015). Increased involuntary musical mental activity is not associated with more accurate voluntary musical imagery.

Psychomusicology: Music, Mind & Brain, 25(1), 48-57. doi:10.1037/pmu0000076

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547-558.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22*(4), 775-789. doi:10.1162/jocn.2009.21239

Chapter 7

General Discussion

"*Mental imagery research has weathered both disbelief of the phenomenon and inherent methodological limitations*" (Pearson et al., 2015, p. 590).

By its nature, mental imagery is private and relatively inaccessible, and requires considerable experimental ingenuity to study. However, research efforts have been richly rewarded with insights into the structure of human cognition, insights bearing on fundamental questions including: whether imagery re-uses the neural machinery of perception (Kosslyn & Thompson, 2003); the nature of mental representations and their transformations (Cooper & Shepard, 1973; Shepard & Metzler, 1971); and whether mental representations of body parts are transformed freely or incorporate knowledge of their biomechanical linkages and constraints (Parsons, 1987; Sekiyama, 1982).

Over the last three decades there has been increased interest in the study of musical imagery from a variety of disciplinary approaches applying a variety of concepts (and confusingly in some cases, a variety of different terms for a given concept) and methodologies (Godøy & Jørgensen, 2001; Schaefer, 2017). This work has shown that many of the features of music (e.g. pitch, rhythm, timbre, loudness) are represented in the musical image (Hubbard, 2013); and that although some studies have shown primary auditory cortex activation (Kraemer et al., 2005), the majority of studies have shown it is the auditory association areas that are most active during imagery, particularly in the right hemisphere (Halpern, 2003; Halpern & Zatorre, 1999; Herholz et al., 2012; Nolden et al., 2013; Otsuka et al., 2008; Zatorre & Halpern, 2005).

The broad aims of this thesis were to (1) develop, test and validate new musical pitch and rhythm imagery tasks that reliably induce musical imagery and are suitable for use in neuroimaging studies; (2) test hypotheses of auditory-sensorimotor brain function during performance of these tasks, using MEG to measure brain function. In the following, we summarise the results of these studies and assess their contributions to the literature.

7.1 Summary of Findings

A review of the literature (Chapter 2) shows that a number of the "inherent methodological limitations" of mental imagery research noted by Pearson et al. (2015) also apply to the musical imagery tasks employed in previous studies. The first study of this thesis (Chapter 3) aimed to develop a new musical pitch imagery task (PIAT) designed to avoid some of these limitations, as follows: first, the PIAT provides objective behavioural confirmation that participants were actually using musical imagery (Hubbard, 2013; Zatorre & Halpern, 2005). Second, each trial is a novel sequence of pitch steps that cannot be anticipated in advance (Herholz et al., 2008; Janata, 2012; Janata & Paroo, 2006). Third, it employs a range of difficulty levels that can accommodate individual differences in musical experience and imagery ability: both musicians and non-musicians are able to complete the task. This is in contrast to previous studies that have either used tasks that are too easy for musicians, such as imaging ascending scales or repeating short patterns of notes (Boh, Herholz, Lappe, & Pantev, 2011; Janata & Paroo, 2006), or too difficult for non-musicians, such as the mental reversal of melodies like Greensleeves (Zatorre et al., 2010). Finally, the imagery task is readily adapted to produce highly comparable control conditions, including music perception and mental arithmetic.

In Chapter 3 we aimed to verify the efficacy of the PIAT in inducing pitch imagery; and to validate it against an established psychometric measure of musical imagery abilities (BAIS). Taken together, the following results provide good evidence that the PIAT is both a selective and valid index of imagery of musical pitch in both musicians and non-musicians: (1) use of an imagery strategy was positively correlated with accuracy of performance on imagery and perception tasks, but was not associated with performance of mental arithmetic; (2) accuracy of imagery was positively correlated with scores on both the vividness and control subscales of the BAIS inventory; (3)

accuracy of performance on the perception task was positively correlated with imagery control but not imagery vividness; (4) accuracy of mental arithmetic was not significantly associated with either of the BAIS scale scores.

Regression analysis confirmed that the most important predictors of PIAT performance were musical imagery strategy use and vividness. While participants with more musical experience did perform better on the PIAT, this relationship was mediated by strategy use. That is, musically trained participants were more likely to use a musical imagery strategy, and therefore do better on the task. Hence Chapter 3 showed that competent performance on the PIAT required an active musical imagery strategy; and that imagery vividness and mental control were more important than musical training for task performance.

Our contention that PIAT indexes musical imagery specifically is further supported by a more detailed examination of the task using Item Response Theory (IRT) (Chapter 4). IRT is a modern psychometric technique used to optimise tests of individual differences. (P. M. C. Harrison et al., 2017). It incorporates two main features in model generation: first, an estimate of the difficulty of a given item based on a generalised mixed effects regression model; second, an estimate of how likely a given participant will complete an item correctly, that is, a model-based ability estimate from the participants. In an exploratory phase a generalised mixed effects regression model was calculated to determine the main predictors of item difficulty. The results showed that performance accuracy was significantly correlated with musical experience and both subscales of the BAIS, but no significant correlations were found between performance and age or gender. The best model of item difficulty found that an item was more difficult if it had (1) higher level (i.e. the number of silent arrows in a trial), (2) lower probability of the probe, (3) the probe note matched the start note of the sequence, (4) higher stages of the task. In

addition the prediction accuracy of the model increased once the random effects of participant ability were added, suggesting that it is a suitable candidate for computerised adaptive testing (CAT) (P. M. C. Harrison et al., 2017). The exploratory phase provided insight into predictors of item difficulty that may be related to perceptual bias (i.e. probability of the probe and probe note matching start note), however due to the staircase design in which higher stages varied in multiple factors it is difficult to isolate the individual contributing factors in order to manipulate. As a result of the exploratory phase, a cognitive process model for the PIAT was developed to describe the steps involved in completing the PIAT. This cognitive process model suggested the need to consider correct and incorrect probes separately in the IRT modelling due to the different cognitive processes that take place.

In the calibration phase, a model was developed that explained performance on the PIAT as the ability to maintain and manipulate pitch representations as well as to resist perceptual biases that can lead to incorrect responses. The advantage of this modelling is that firstly, it provides an opportunity to quantify the effect of perceptual bias in the task, an area of music cognition which is often difficult to study systematically; secondly, while internal mental manipulation of sounds or musical elements is a core component of Gordon's audiation concept, tests of audiation to date have not required such manipulation to take place (Gordon, 1989a). The results of Chapter 4 show that the PIAT is an effective index of Gordon's conceptualisation of audiation.

Previous neuroimaging and neurophysiological studies (Chapter 2) have shown that brain networks connecting auditory and motor processing centres are importantly involved in the mediation of music perception. The MEG studies of Chapters 5 and 6 were designed to examine the activities and connectivity of these brain regions during mental imagery of musical pitch and rhythm, respectively.

The work described in Chapter 5 used the PIAT to induce musical pitch imagery while beta-band activity in auditory and sensorimotor cortical sources were measured with MEG. The spatial sensitivity of the sources was confirmed as the bilateral auditory sources were maximally sensitive to the onset of tones, whilst the left sensorimotor source showed a typical readiness field in anticipation of hearing the probe and making a response with the right index finger. The results showed (1) greater beta-band modulation during imagery than perception; (2) no regional or hemispheric differences in beta-band modulation; (3) connectivity directed from auditory to sensorimotor cortex during perception, but the opposite during imagery (sensorimotor to auditory); (4) the degree of sensorimotor to auditory connectivity in anticipation of a tone in perception was predicted by imagery accuracy.

Beta oscillations have been argued to track both the 'what' and the 'when' of auditory events. Beta-ERS has been related to anticipated timing of future events (Fujioka et al., 2012) whilst beta-ERD has been related to the predictability of what a future note may be (Chang et al., 2018), as well as marking a salient event that has just occurred ("event tagging") (Hanslmayr & Staudigl, 2014; Teki & Kononowicz, 2016). The results of Chapter 5 show greater beta modulation during pitch imagery in a task that crucially has a constant presentation rate. Hence if beta modulation is related to event tagging this suggests that beta-band is involved not only in 'when' but also 'what' is to occur, even when this event is an internal generation or manipulation of a sound image. Alternatively, if beta modulation is related to predictive coding, then the greater modulation in the imagery condition could be due to the predictability of the event following. That is, once the first silent arrow appeared in a trial, participants would have known there were 2 more silent arrows and a pre-probe screen to follow in the imagery trial. However, in the perception trial throughout the final 3 arrows participants were not aware they were

completing a perception trial and were most likely actively preparing in case a silent arrow was about to appear. Thereby making the perception trials less predictable than the imagery trials. Future studies could clarify which interpretation is correct by indicating when the last 3 arrows were occurring in both trials (i.e. through colour change of arrow), and therefore making trials equal in predictability. If imagery trials still show greater beta modulation than perception this would be evidence for an event tagging mechanism.

The similar time courses of the beta-band in the auditory and sensorimotor regions in all three conditions is consistent with the notion of "communication through coherence" (Fries, 2015); that the beta-band is an "open line of communication" between the regions (Tang et al., 2016), and a mechanism that supports auditory-motor connectivity (Henry & Grahn, 2017). Such coordinated activity is predicted by forward and inverse feedback loops between these regions in the dorsal pathway (Rauschecker & Scott, 2009).

Connectivity analysis showed a left hemisphere dominance during mental manipulation of pitches, consistent with the results of brain stimulation studies showing stimulation to the left supramarginal gyrus enhances pitch memory but not rhythm memory (Schaal et al., 2017) and TMS at the theta-band frequency to the left intraparietal sulcus enhances melody manipulation but not melody maintenance (Albouy et al., 2017). Further, the positive association found between pitch imagery accuracy and the degree of left sensorimotor to right auditory region directed connectivity during the 666 – 999 ms epoch after a sounded arrow (i.e. the 333 ms epoch before the next arrow) shows the first evidence for an association between temporal prediction in perception and imagery ability. While it has been argued that imagery is involved in perception, in what Janata (2001b) refers to as "expectant imagery" or Moore (2011) and Schaefer (2017) refer to as

"constructive imagery", no clear empirical evidence has been found for such a link between explicit musical imagery and implicit processes involved in music perception.

The PIAT requires mental imagery of changing pitch but maintains a constant timing or rhythm. Chapter 6 describes a converse task with constant pitch and variable rhythm, in a maintenance paradigm: the Rhythm Imagery Task (RIT). In an initial behavioural study, results showed that performance (both accuracy and reaction times) in imagining rhythms was significantly improved after a block of tapping out the patterns, even when the auditory consequences of the taps were masked by white noise. Imagery accuracy and tapping accuracy were both positively correlated with Musical Training and General Sophistication subscales of the Gold-MSI, whilst no significant correlations were found with either BAIS subscale. In addition, better performance on tapping trials was associated with lower variability in tapping during the syncopated pattern. Multiple regression analysis revealed that tapping accuracy and reaction time in the RIT after tapping were the most significant predictors of RIT accuracy in Block 2.

These findings are consistent with previous movement and no-movement (i.e. imagery) isochronous rhythm studies that have shown that moving to the beat leads to greater temporal precision than no-movement (Manning & Schutz, 2013, 2015), and that in non-percussionists, less variability in tapping was associated with greater accuracy in the movement conditions (Manning et al., 2017; Manning & Schutz, 2016). As tapping accuracy was the most important factor in subsequent rhythm imagery performance, this supports the interpretation that the brain's motor systems are involved in rhythm imagery (Manning et al., 2017; Manning & Schutz, 2015).

The MEG study in Chapter 6 confirmed that RIT performance did improve after tapping, but only as seen in faster reaction times, with accuracy remaining relatively high across both imagery blocks. This may have been due to a ceiling effect. No significant

correlations between performance measures and musical training, BAIS subscales or tapping measures were found. The main results of the MEG study were (1) greater beta-band modulation of sensorimotor regions than the auditory regions at the bass drum frequency during perception and imagery, particularly in the right sensorimotor region; (2) no differences between sensorimotor and auditory regions at the beat frequency (2.5 Hz) during perception or imagery; (3) greater beta-band modulation at the beat frequency (2.5 Hz) in perception than imagery; (4) effect of tapping was seen most strongly in changes in directed connectivity in theta-band from the right sensorimotor region for both perception and imagery; (5) rhythm imagery accuracy was associated with greater right sensorimotor to left auditory driven connectivity after tapping during the strong beat rest in the syncopated pattern.

Taken together these results point to the importance of the right sensorimotor region in both perceiving and imagining rhythmic patterns, suggesting that the region plays a leading role at points of metrical importance, namely the bass drum frequency (0.625 Hz in the beta-band modulation) and the strong 3rd beat of the bar (800 – 1200 ms in the theta band dPTE analysis). These results are consistent with the proposal that motor regions are involved in beat perception (Grahn & Brett, 2007) and that rhythmic stimulation by bass sounds leads to enhanced neural representation of the beat and meter (Lenc et al., 2018). Previously the right auditory cortex has been implicated in tracking the beat in syncopated rhythms in which there is no energy at the beat frequency (Tal et al., 2017), however in that task participants were passively listening to the patterns and were not instructed to maintain the beat. Given the need to actively maintain the rhythmic pattern in silence, our results may have resulted in greater recruitment of the sensorimotor region, however the right hemisphere dominance was consistent. In addition, brain

stimulation studies have shown stimulation to the right supramarginal gyrus leads to enhanced rhythm memory capacity (Schaal et al., 2017).

7.2 Contributions to the literature

The results of this thesis make the following methodological and empirical contributions the musical imagery literature.

7.2.1 Validated Test of Pitch Imagery

To date, limited tools exist for researchers to approximate the pitch imagery ability of participants. While melodic discrimination tasks have been used to obtain a measure of audiation ability (Gordon, 1989a), and pitch memory spans, equivalent to digit spans, have been used to determine pitch working memory capacity (Schaal et al., 2014; Schaal et al., 2013), neither of these tasks provide a measure of active musical imagery in which participants must both maintain and manipulate the sound image. In addition, the literature review showed several short comings of other pitch imagery tasks that have been previously used (Chapter 2). The outcome of this thesis is an efficient measure of pitch imagery ability in an adaptive online task that takes just 10 minutes to complete. Ability on this task involves both the ability to maintain and manipulate pitch images, as well as the ability to overcome perceptual biases; an aspect of music cognition that up until now has been difficult to measure objectively. A demonstration of the updated PIAT is available online (http://shiny.pmcharrison.com/piat-demo/) and future studies are underway to validate this task against established measures of melodic discrimination (P. M. C. Harrison et al., 2016), visuo-spatial working memory (Vock & Holling, 2008), and auditory working memory tasks such as the backwards digit span (Wechsler, 2008). The ultimate goal is to establish the task as a simple tool for other researchers in the field to use as a validated measure of pitch imagery (see for example, Colley et al., 2018).

7.2.2 Auditory-Sensorimotor Interactions

The primary aim of this thesis was to investigate auditory-sensorimotor brain function during performance of pitch and rhythm imagery, using MEG. While auditory-sensorimotor interactions have recently been explored in the music perception literature (Morillon & Baillet, 2017; Ross et al., 2017; Tal et al., 2017), the MEG studies discussed in Chapter 5 and 6 were the first to examine the oscillatory mechanisms at work in these regions during silence as participants imagined pitches and rhythms. Both studies showed that during imagery, beta rhythms in the auditory and sensorimotor regions were modulated in a similar fashion, and that imagery accuracy is related to greater top-down (sensorimotor to auditory) directed connectivity during silences in perception. Hemispheric differences were found between pitch and rhythm imagery, as well as differences in the relative amount of beta-band modulation when compared to perception.

The similar beta-band modulation seen in auditory and sensorimotor regions during imagery of pitch manipulation and rhythm maintenance is consistent with a mechanism supporting auditory-motor connectivity (Henry & Grahn, 2017), and an "open-line" of communication between the regions (Tang et al., 2016). This has been previously shown in passive listening (Fujioka et al., 2012), with the right postcentral region showing a similar time course of beta coherence to the right Heschl's gyrus, however this is the first evidence for co-modulation during imagery in silence. This contributes new evidence to the field: that the auditory-sensorimotor regions are coordinating their activities during internal generation of music in the mind. While this effect in the pitch imagery study could be argued to be in response to the visual onset of the arrow at 1 Hz, in the rhythm imagery study the only external stimuli was the bass drum beat at 0.625 Hz.

Both pitch and rhythm imagery accuracy were related to increased sensorimotor to auditory directed connectivity during silences in pitch and rhythm perception respectively. The notion that musical imagery can be conceived as both a voluntary mental act as well as an implicit act required for music perception is not new (Janata, 2001b; Schaefer, 2017), however to date limited studies have shown evidence for a relationship between imagery ability and neural activation during music perception. One study that has looked at structural correlates of subjective vividness of imagery found a significant positive relationship between BAIS-V scores and grey matter volume in a cluster centred in the left SMA and extending to the left and right paracentral lobules (Lima et al., 2015). In addition, that study had participants passively listen to 5 different types of human vocal sounds and found the representational similarity of these sounds in the SMA was negatively associated with BAIS-V scores, indicating that participants with greater vividness had more distinct representations of different auditory sounds (Lima et al., 2015). In a study of verbal imagery, activity in the SMA preceded the auditory regions during imagery of speech, however in auditory hallucinations such directed activity was not seen, with the two regions activating at the same time (Linden et al., 2011). Hence this thesis has shown for the first time, an association between objective imagery ability and increased sensorimotor to auditory directed connectivity during music perception.

Using the same MEG analysis pipeline, the results of this thesis showed hemispheric differences in neurophysiological activities during pitch and rhythm imagery. The literature (Chapter 2) had previously shown that pitch and rhythm are dissociable, as evidenced by behavioural tasks where temporal imagery was more difficult that pitch imagery (Janata & Paroo, 2006; Weir et al., 2015) and by brain stimulation studies showing left hemisphere effects on pitch memory and right hemisphere effects on rhythm memory (Schaal et al., 2017). Left hemisphere brain

stimulation also improved performance on manipulation (i.e. reversal) but not maintenance of short melodies (Albouy et al., 2017). The current results showing an association between pitch imagery ability and left sensorimotor to right auditory directed connectivity is consistent with the results of both brain stimulation studies. Conversely, the association between right sensorimotor to left auditory directed connectivity during the strong beat of the rhythmic patterns after tapping is consistent with the role of the right hemisphere in rhythm memory.

Finally, while the amount of beta-band modulation during pitch imagery was greater than pitch perception (Chapter 5), this was not true for rhythm imagery, where the beta-band modulation at the beat rate was greater during heard patterns than imagined patterns (Chapter 6). This may have been due to the temporal nature of the rhythm imagery task, where imagined tones were not as consistently in time as heard tones. It may also be due to the predictability of temporal events in each task.

7.2.3 Individual Differences

This thesis addressed individual differences in musical training, as well as psychometric measures of imagery ability and compared them with performance on pitch and rhythm imagery tasks. Musical training has been shown to improve both pitch and temporal acuity of imagery (Aleman et al., 2000; Janata & Paroo, 2006). This thesis has shown that performance on the PIAT (Chapter 3) was best predicted by the variables of strategy use (musical imagery vs alternative strategy) and vividness, and that vividness and control were more important than musical training to be able to do the task. With larger numbers of participants (Chapter 4), a positive association was found between PIAT performance and musical training, vividness as well as mental control. In the behavioural study of the rhythm imagery task (Chapter 6) musical training and general sophistication were significantly correlated with accuracy of both rhythm imagery and

tapping, whereas vividness and control were not significantly related to performance. In contrast to the PIAT, the best predictors of performance accuracy on the RIT were higher tapping accuracy and faster reaction times on RIT. Taken together these results suggest auditory imagery vividness and control may be more important for pitch imagery than for rhythm imagery, and that musical training is associated with better pitch and rhythm imagery performance but are not the most important predictors of performance. This is consistent with previous research that has shown musicians can recall longer sequences of spoken digits than non-musicians, but perform no better at a backward span task in which the verbal image must be manipulated (Hansen et al., 2013). Hence the finding that musical training was more associated with RIT performance than PIAT performance may be due to the fact that mental manipulation was required to successfully complete the PIAT, where mental maintenance was employed for the RIT.

7.2.4 Movement and Musical Imagery

The findings of Chapter 6 confirmed that short-term motor engagement improves subsequent rhythm imagery performance, and that the right sensorimotor region plays a role in tracking the downbeats (and strong beats) of both heard and imagined rhythmic patterns. This overlap in processing of the beat could be one potential explanation for the overlap in processing seen between perception and imagery in general (Hubbard, 2013; Schaefer, 2017). While previous studies have compared movement and no-movement conditions during temporal prediction and beat perception, these studies have either interspersed conditions (Manning et al., 2017; Manning & Schutz, 2013, 2015, 2016; Morillon & Baillet, 2017), or if a block design was used, did not compare performance in the movement first and no-movement first conditions (Stupacher et al., 2016). Hence these designs do not comprehensively answer the question of the role of movement in temporal prediction or beat perception. By using an Imagery-Tapping-Imagery design in

chapter 6, this thesis overcomes this issue and presents the first study to show an immediate effect of short-term tapping training on rhythm imagery performance, as well as providing evidence for the neuroplastic changes that occur in connectivity from the right sensorimotor region as a consequence of tapping.

7.3 Conclusions

The aim of this thesis was to develop, test and validate new musical imagery tasks; and to use them to test hypotheses of auditory-sensorimotor brain function during mental imagery of pitch and rhythm. The results support the following conclusions: auditory-sensorimotor brain regions coordinate their activities during pitch and rhythm imagery; individual differences in imagery ability are related to perceptual processing; imagery vividness and mental control are more important for pitch imagery than for rhythm imagery; musical training is associated with better imagery performance but is not the most important factor for either pitch or rhythm imagery performance; and short-term motor training improves subsequent rhythm imagery performance and modulates the amount of right sensorimotor activity in both rhythm imagery and perception.

The results of the thesis provide new tools for research, and the PIAT has recently been used by other investigators including Colley et al. (2018). These new tools may also have utility in clinical settings. One growing area of applied research is the use of music education for cochlear implant recipients, for improving perceptual outcomes and music appreciation (Lo, McMahon, Looi, & Thompson, 2015). A simplified version of the PIAT / RIT may be of utility in this setting by virtue of their features including: computerised adaptive design allowing for efficient testing of participants with a range of ability levels from very simple to challenging, validation against established psychometric measures of musical ability, and objective behavioural measures of performance.

The results of the thesis also contribute to our understanding of the neurophysiological mechanisms that support our ability to imagine music. Taken together, the MEG results indicate that representations and transformations of musical images draw on the same auditory-sensorimotor brain networks that have been suggested to support music perception (Krishnan et al., 2018; Morillon & Baillet, 2017; Ross et al., 2017; Tal et al., 2017). The notion that imagery and perception draw on common sets of neural machinery is entirely consistent with a consensus emerging from results of functional imaging studies in the visual modality (Pearson et al., 2015).

7.4 References

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron*, 94(1), 193-206. doi:10.1016/j.neuron.2017.03.015

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia*, 38(12), 1664-1668. doi:10.1016/S0028-3932(00)00079-8

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PLoS ONE*, 6(7), e21458. doi:10.1371/journal.pone.0021458

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2018). Beta oscillatory power modulation reflects the predictability of pitch change. *Cortex*, 106. doi:10.1016/j.cortex.2018.06.008

Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronization with expressively timed music. *The Quarterly Journal of Experimental Psychology*, 71(8), 1781 - 1796. doi:10.1080/17470218.2017.1366531

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1), 220-235. doi:http://dx.doi.org/10.1016/j.neuron.2015.09.034

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *The Journal of Neuroscience*, 32(5), 1791-1802. doi:10.1523/jneurosci.4107-11.2012

Godøy, R. I., & Jørgensen, H. (2001). *Musical imagery*. Lisse, the Netherlands: Swets & Zeitlinger Publishers.

Gordon, E. E. (1989). *Advanced measures of music audiation*. Chicago: Riverside Publishing Company.

Grahn, J. A., & Brett, M. (2007). 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

Halpern, A. R. (2003). Cerebral substrates of musical imagery. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 217-230). New York, NY: Oxford University Press; US.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9(7), 697-704. doi:10.1093/cercor/9.7.697

Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, 41(6), 779-793. doi:10.1177/0305735612452186

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories — A processing based perspective on oscillatory subsequent memory effects. *NeuroImage*, 85, Part 2, 648-655. doi:10.1016/j.neuroimage.2013.05.121

Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7(1), 3618. doi:10.1038/s41598-017-03586-z

Harrison, P. M. C., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: Descriptive and explanatory approaches. *Journal of New Music Research*, 45(3), 265-280. doi:10.1080/09298215.2016.1197953

Henry, M. J., & Grahn, J. A. (2017). Music, brain, and movement. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition*. Abingdon: Routledge.

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, 24(6), 1382-1397. doi:10.1162/jocn_a_00216

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience*, 28(11), 2352-2360. doi:10.1111/j.1460-9568.2008.06515.x

Hubbard, T. L. (2013). Auditory aspects of auditory imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory Imagery* (pp. 51 - 76). S.I.: Springer New York.

Janata, P. (2001). Neurophysiological mechanisms underlying auditory image formation in music. In R. I. Godøy & H. Jørgensen (Eds.), *Musical Imagery*. New York: Routledge.

Janata, P. (2012). Acuity of mental representations of pitch. *Annals of the New York Academy of Sciences*, 1252(1), 214-221. doi:10.1111/j.1749-6632.2011.06441.x

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception & Psychophysics*, 68(5), 829-844. doi:10.3758/BF03193705

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129(5), 723-746.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature*, 434(7030), 158.

Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1801421115

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., . . . Scott, S. K. (2015). Feel the noise: relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex*, 25(11), 4638-4650. doi:10.1093/cercor/bhv134

Linden, D. E. J., Thornton, K., Kuswanto, C. N., Johnston, S. J., van de Ven, V., & Jackson, M. C. (2011). The brain's voices: Comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex*, 21(2), 330-337. doi:10.1093/cercor/bhq097

Manning, F. C., Harris, J., & Schutz, M. (2017). Temporal prediction abilities are mediated by motor effector and rhythmic expertise. *Experimental Brain Research*, 235(3), 861-871. doi:10.1007/s00221-016-4845-8

Manning, F. C., & Schutz, M. (2013). "Moving to the beat" improves timing perception. *Psychonomic bulletin & review*, 20(6), 1133-1139. doi:10.3758/s13423-013-0439-7

Manning, F. C., & Schutz, M. (2015). Movement enhances perceived timing in the absence of auditory feedback. *Timing & Time Perception*, 3(1-2), 3-12. doi:10.1163/22134468-03002037

Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research*, 80(4), 532-542. doi:10.1007/s00426-015-0678-5

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences*, 114(42), E8913-E8921. doi:10.1073/pnas.1705373114

Nolden, S., Grimault, S., Guimond, S., Lefebvre, C., Bermudez, P., & Jolicoeur, P. (2013). The retention of simultaneous tones in auditory short-term memory: A magnetoencephalography study. *NeuroImage*, 82, 384-392. doi:10.1016/j.neuroimage.2013.06.002

Otsuka, A., Tamaki, Y., & Kuriki, S. (2008). Neuromagnetic responses in silence after musical chord sequences. *NeuroReport*, 19(16), 1637-1641. doi:10.1097/WNR.0b013e32831576fa

Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, 19(2), 178-241. doi:10.1016/0010-0285(87)90011-9

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, 19(10), 590-602. doi:10.1016/j.tics.2015.08.003

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724. doi:10.1038/nn.2331

Ross, B., Barat, M., & Fujioka, T. (2017). Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced beta-band oscillations during perception. *The Journal of Neuroscience*, 37(24), 5948-5959. doi:10.1523/jneurosci.3613-16.2017

Schaal, N. K., Krause, V., Lange, K., Banissy, M. J., Williamson, V. J., & Pollok, B. (2014). Pitch memory in nonmusicians and musicians: Revealing functional

differences using transcranial direct current stimulation. *Cerebral Cortex*. doi:10.1093/cercor/bhu075

Schaal, N. K., Pollok, B., & Banissy, M. J. (2017). Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Scientific Reports*, 7, 42456. doi:10.1038/srep42456

Schaal, N. K., Williamson, V. J., & Banissy, M. J. (2013). Anodal transcranial direct current stimulation over the supramarginal gyrus facilitates pitch memory. *European Journal of Neuroscience*, 38(10), 3513-3518. doi:doi:10.1111/ejn.12344

Schaefer, R. S. (2017). Music in the brain: imagery and memory. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition*. Abingdon: Routledge.

Sekiyama, K. (1982). Kinesthetic aspects of mental representations in the identification of left and right hands. *Perception & Psychophysics*, 32(2), 89-95. doi:10.3758/BF03204268

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701-703. doi:10.1126/science.171.3972.701

Stupacher, J., Witte, M., Hove, M. J., & Wood, G. (2016). Neural entrainment in drum rhythms with silent breaks: Evidence from steady-state evoked and event-related potentials. *Journal of Cognitive Neuroscience*, 28(12), 1865-1877. doi:10.1162/jocn_a_01013

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017). Neural entrainment to the beat: the “missing-pulse” phenomenon. *The Journal of Neuroscience*, 37(26), 6331-6341. doi:10.1523/jneurosci.2500-16.2017

Tang, H., Crain, S., & Johnson, B. W. (2016). Dual temporal encoding mechanisms in human auditory cortex: Evidence from MEG and EEG. *NeuroImage*, 128, 32-43. doi:10.1016/j.neuroimage.2015.12.053

Teki, S., & Kononowicz, T. W. (2016). Commentary: Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. *Frontiers in Neuroscience*, 10(389). doi:10.3389/fnins.2016.00389

Vock, M., & Holling, H. (2008). The measurement of visuo–spatial and verbal–numerical working memory: Development of IRT-based scales. *Intelligence*, 36(2), 161-182. doi:10.1016/j.intell.2007.02.004

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th Edition ed.). San Antonio, TX: Pearson.

Weir, G., Williamson, V. J., & Mullensiefen, D. (2015). Increased involuntary musical mental activity is not associated with more accurate voluntary musical imagery. *Psychomusicology: Music, Mind & Brain*, 25(1), 48-57. doi:10.1037/pmu0000076

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron*, 47(1), 9-12. doi:10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, 22(4), 775-789. doi:10.1162/jocn.2009.21239

MACQUARIE University

REBECCA GELDING
<rebecca.gelding@students.mq.edu.au>

Approved - HREC App REF 5201300453

3 messages

Ethics Secretariat <ethics.secretariat@mq.edu.au> Fri, Jul 26, 2013 at 8:58 AM
To: Dr Blake Johnson <blake.johnson@mq.edu.au>
Cc: Mrs Rebecca Wynne Gelding <rebecca.gelding@students.mq.edu.au>

Dear Associate Professor Johnson

RE: "Behavioural and MEG Study of Musical Imagery and Perception in Adults" (REF: 5201300453)

Thank you for your correspondence dated 22 July 2013 responding to the issues raised by the Macquarie University Human Research Ethics Committee (HREC (Medical Sciences)). The HREC (Medical Sciences) delegated the review of your responses to the Ethics Secretariat.

The HREC (Medical Sciences) is fully constituted and operates in accordance with the National Health and Medical Research Council's National Statement on Ethical Conduct in Human Research (2007) (the National Statement) and the CPMP/ICH Note for Guidance on Good Clinical Practice.

I am pleased to advise that the above project has been granted ethical and scientific approval.

This research meets the requirements of the National Statement which is available at the following website:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf

This letter constitutes ethical and scientific approval only.

The following documentation has been reviewed and approved by the HREC (Medical Sciences):

1. Macquarie University Ethics Application Form (v 2.1 - Feb 2013)
2. Macquarie University Participant Information and Consent Form "Behavioural and MEG Study of Musical Imagery and Perception in Adults" (v 2, July 2013)
3. Advertisement (undated)
4. Music Education and preference questionnaire incl:
a. Bucknell Auditory Imagery Vividness Scale
b. Bucknell Auditory Imagery Control Scale

c. Debrief Questions

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement. It is the responsibility of the Principal Investigator to ensure that the protocol complies with the HREC-approval and that a copy of this letter is forwarded to all project personnel.

2. The National Statement sets out that researchers have a "significant responsibility in monitoring, as they are in the best position to observe any adverse events or unexpected outcomes. They should report such events or outcomes promptly to the relevant institution/s and ethical review body/ies, and take prompt steps to deal with any unexpected risks" (5.5.3).

Please notify the Committee within 72 hours of any serious adverse events or Suspected Unexpected Serious Adverse Reactions or of any unforeseen events that affect the continued ethical acceptability of the project.

3. Approval will be for a period of five (5) years subject to the provision of annual reports.

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.
Progress reports and Final Reports are available at the following website:
http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

4. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

5. All amendments to the project must be reviewed and approved by the Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the Hospital and University. This information is available at the following websites:

http://www.mq.edu.au/policy/

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have ethics approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.
If you need to provide a hard copy letter of ethics approval to an external organisation as evidence that you have approval please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of ethics approval.

Yours sincerely
Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committee (Medical Sciences)

Ethics Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 2 9850 6848
F: +61 2 9850 4465
http://www.mq.edu.au/research

CRICOS Provider Number 00002J

Please consider the environment before printing this email.
This email (including all attachments) is confidential. It may be subject to legal professional privilege and/or protected by copyright. If you receive it in error do not use it or disclose it, notify the sender immediately, delete it from your system and destroy any copies. The University does not guarantee that any email or attachment is secure or free from viruses or other defects. The University is not responsible for emails that are personal or unrelated to the University's functions.