The Behavioural and Neural Bases of Music-Evoked Visual Mental Imagery

Sarah Said Hashim

March 2024

Department of Psychology

Goldsmiths, University of London

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctorate of Philosophy in Psychology

Statement of Originality

I, Sarah Said Hashim, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

All work was carried out by myself except for the cases presented below:

- I would like to formally acknowledge and thank Olivia Geibel for her assistance with coding the visual imagery content descriptions during the thematic analysis for the research presented in Chapter 2.
- Conception of the design and all data collection of the research presented in Chapter 3 were carried out by Prof André Weinreich and Dr Mats B. Küssner.
- The Ambient track used in the research presented in Chapters 4 and 5 was provided by the sonic branding agency, MassiveMusic, as part of a collaboration.
- The pool of aphantasic participants contacted in both online studies in the research presented in Chapter 6 was compiled by fellow PhD student, Claudia Pulcini.

Signed: Date: *28th March 2024*

Acknowledgements

I am immensely grateful to my exceptional supervisors, Dr Diana Omigie and Dr Mats B. Küssner, whose guidance, brilliance, and encouragement have been invaluable throughout my PhD journey Their belief in me has fuelled my confidence, and I am truly honoured to have learned from such inspiring mentors. I extend my thanks to Prof Laurent Stewart for her continuous interest in my work. I could not have asked for more inspiring role models.

To my officemates, for the friendship, bond, and laughs that we shared together. I would like to thank all the participants who took part in my research, without whom this work would not have been possible.

My deepest appreciation goes to my family for their unwavering love and support during the times I needed it most. I thank my incredible friends for giving me strength on this journey. Finally, thank you to Yusuf for cheering me on until the end.

To each and every individual who has played a part, in shaping my academic and personal growth, I extend my deepest gratitude. Your contributions have left a lasting mark on my journey, and for that, I am truly thankful.

Related Publications and Presentations

A version of the research reported in Chapter 2 is published in:

Hashim, S., Stewart, L., Küssner, M. B., & Omigie, D. (2023). Music listening evokes story-like visual imagery with both idiosyncratic and shared content. *PLOS ONE*, *18*(10), e0293412. https://doi.org/10.1371/journal.pone.0293412

A version of the research reported in Chapter 3 is published in:

Hashim, S., Küssner, M. B., Weinreich, A., & Omigie, D. (2024). The neuro-oscillatory profiles of static and dynamic music-induced visual imagery. *International Journal of Psychophysiology*, *199*, 112309. https://doi.org/10.1016/j.ijpsycho.2024.112309

A version of the research reported in Chapter 6 is published in:

Hashim, S., Pulcini, C., Jansari, A., Küssner, M. B., & Omigie, D. (2024). The Experience of Music in Aphantasia: Emotion, Reward, and Everyday Functions. *Music & Science*, *7*, 20592043231216259. https://doi.org/10.1177/20592043231216259

Additional publications related to the present work include:

Antović, M., Küssner, M. B., Kempf, A., Omigie, D., **Hashim, S.**, & Schiavio, A. (2024). “A huge man is bursting out of a rock”: bodies, motion, and creativity in verbal reports of musical connotation. *Journal of New Music Research*, *52*(1), 73–86. https://doi.org/10.1080/09298215.2024.2306406

The work presented in Chapters 2 – 6 have been presented at various conferences:

Hashim, S., & Omigie, D. (2023b). Music Listening and Stress Reduction: Visual Imagery as an Underlying Mechanism? [Poster Presentation]. *SysMus23*, Sheffield, United Kingdom.

Hashim, S., & Omigie, D. (2023a). A probe-caught neurophysiological investigation of visual imagery during extended listening [Oral Presentation]. *SysMus23*, Sheffield, United Kingdom.

Hashim, S., & Omigie, D. (2022). A Probe-Caught Investigation of Spontaneous & Deliberate Visual Imagery [Oral Presentation]. *Sempre 50th Anniversary Conference*, London, United Kingdom.

Hashim, S., Küssner, M. B., Weinreich, A., & Omigie, D. (2021e). Predicting Music-Induced VisualImagery using Occipital Alpha [Online Oral Presentation]. *ICMPC16-ESCOM11*.

Hashim, S., Stewart, L., Küssner, M. B., & Omigie, D. (2021d). Evaluating the Consistency and Thematic Content of Music-Induced Visual Mental Imagery [Online Oral Presentation]. *ICMPC16-ESCOM11*.

Hashim, S., Stewart, L., Küssner, M. B., & Omigie, D. (2021c). Evaluating the consistency and thematic content of music-induced visual mental imagery [Poster Presentation]. *The Neurosciences and Music VII*, Ontario, Canada.

Hashim, S., Küssner, M. B., Weinreich, A., & Omigie, D. (2021b). Predicting Music-Induced Visual Imagery Using Occipital Alpha [Online Poster Presentation]. *8th MindBrainBody Symposium*.

Hashim, S., Stewart, L., Küssner, M. B., & Omigie, D. (2021a). Evaluating the consistency and thematic content of music-induced visual mental imagery [Spoken Presentation]. In J. Stupacher & S. Hagner (Eds.), *Proceedings of the 14th International Conference* *of Students of Systematic Musicology*. Open Science Framework. https://doi.org/10.17605/OSF.IO/HG6RZ

Hashim, S., Küssner, M. B., Stewart, L., & Omigie, D. (2020b). Evaluating the Consistency and Thematic Content of Music-Induced Visual Mental Imagery [Poster Presentation]. *16th Annual NeuroMusic Conference*. Music in Cross-Cultural Perspective, Ontario, Canada [Online].

Hashim, S., Stewart, L., Küssner, M. B., & Omigie, D. (2020a). The Relationship Between General Visual Imagery and Music-Induced Visual Imagery [Online Poster Presentation]. *German Society for Music Psychology*.

Table of Contents

List of Figures

List of Tables

Abstract

Visual mental imagery, the experience of "seeing" in the mind's eye in the absence of the corresponding retinal input, is a ubiquitous experience that holds artistic, clinical, and educational implications. The acoustic and associative features of music make it an effective trigger of visual imagery. However, while visual imagery has been extensively explored across several disciplines, there is a lack of basic knowledge and methodological breadth in investigations of music-evoked visual imagery. Consequently, there is much to be understood regarding its phenomenological features, its neural characteristics, and its practical impact. This thesis seeks to address these outstanding gaps across five empirical investigations. In the first experiment, qualitative and quantitative techniques are used to examine the content and the across- and within-person consistency of music-evoked visual imagery. Secondly, electroencephalography (EEG) was used to explore the neural substrates of static and dynamic forms of music-evoked visual imagery. Thirdly, EEG was used again to understand and compare the neuro-oscillatory characteristics of spontaneously- and deliberately-occurring visual imagery during extended music listening. Next, using self-reports, physiology, and EEG, visual imagery was assessed as a potential mechanism for stress reduction following a multicomponent stress-induction task. In the final experiment, the thesis explored whether aphantasics, those with little-to-no visual imagery ability, differ in their emotional, functional, and everyday uses of music compared to control listeners. By combining a variety of measurement and analytical techniques, this thesis is able to emphasise the diverse, idiosyncratic, and spontaneous nature of music-evoked visual imagery, reveal its neural characteristics, and demonstrate its ability to modulate affect.

CHAPTER 1. THEORETICAL BACKGROUND

This chapter provides an introduction into the existing literature on music-evoked visual mental imagery. It will situate visual imagery within the wider literature of imagination, before introducing its role in music listening by drawing upon research conducted within the cognitive and neuroscientific domains over the last few decades. Important findings gained thus far as well as apparent gaps are highlighted throughout. The chapter concludes by presenting the research questions that the thesis broadly aims to address and by providing a brief outline of the empirical chapters that will follow.

1.1. INTRODUCTION TO CORE CONCEPTS

1.1.1. Imagination

Imagination is crucial in the creation of fictional stories (Margulis & Jakubowski, 2024), as well as in the construction and reconstruction of personal events and possible future scenarios (Schacter & Addis, 2007). It typically entails the conjuration of a conceptual space that fosters different modes of representation, and is considered a way of conceiving alternate realities, problem-solving, and of propelling creativity (Abraham, 2020). Imagination represents a departure from the here-and-now and, as a constructive process, relies on memory to formulate mental scenarios where fact and fiction intermingle (Hassabis et al., 2007; Moulton & Kosslyn, 2009).

Some authors argue that imagination emanates from semantic memory, which creates a basis and forms a pathway for the brain to make connections between semantic, spatial, and episodic information (Abraham & Bubic, 2015). Other authors hypothesise that imagination is an adaptive process, that forms an amalgamation of raw materials from episodic memory that are flexibly constructed to form simulations of future events, known as the constructive episodic simulation hypothesis (Gaesser, 2013; Schacter & Addis, 2007). Bridging these two theories together, Devitt et al. (2017) argued that while the roles of semantic and episodic memory are dissociable processes, they form a co-dependent compensatory relationship whereby semantic memory aids in embellishing episodically low content.

Imagination possesses general implications in everyday life. Visual imagination is understood to enrich creative experiences in the arts, essentially forming a cognitive canvas against which an individual re-evaluates and reforms lived experiences (Aldworth, 2018). Ultimately, different elements of imagination are represented by different sensory information, especially within the visual domain. In the case of memory representation, D'Argembeau and Van der Linden (2004) explored the phenomenal characteristics of mentally “re-experienced” or “pre-experienced” positive and negative events, finding that individuals tend to adopt visual first- or third-person perspectives of themselves that were in turn related to the reported temporal distance of an event. Further, Schacter and Addis (2007) found activity in posterior regions of the brain, responsible for processing visual information, in the construction of past and future events, providing neural evidence that recollections of memories are represented visually.

In fact, in a framework formulated by Abraham (2016) to bridge the gap between behavioural and neuroscientific discourses and taxonomies of imagination, “mental imagery”, including visual, auditory, motor, and other sensory domains, is cited to be a common category of information processing and, as such, a core classification of imagination. Additionally, “music-related aesthetic response” (Abraham, 2016, p. 4201) is proposed to be an emotional phenomenological proponent of imagination (amongst others, e.g., aesthetic engagement, literature-related aesthetic response), due to music representing an evocative artform.

Perspectives on the role of mental imagery in imagination differ across disciplines and highlight various points of departure. According to Nanay (2021), the involuntary nature of mental imagery does not sufficiently classify it as a category of imagination, as imagination is inherently a voluntary process. Perspectives from philosophy question what is contained in imagination, and propose that mental imagery is not essential for imagination, but that it plays a role in the contents of imaginative episodes (Gregory, 2016). Others suggest that one needs to widen the scope of mental imagery, and consider that mental imagery serves different imaginative purposes that are dependent on its content and its attitude (Arcangeli, 2020). In any case, common threads in existing research are that mental imagery constitutes a type of "sensory imagination", and that while imagery and imagination can exist independently, they also seamlessly co-exist and share phenomenological overlap. In this thesis, I will speak specifically of *visual* imagery, which I will characterise, with regard to various levels of inquiry, as a creative, automatic, and beneficial outlet of the imaginative human mind.

1.1.2. Visual mental imagery

Reports on the details into the phenomenal features of visual mental imagery date back to the start of the 20th century. A critical debate developed in the 1970's amongst the domains of philosophy and cognitive psychology and concerned the format of the mental representation of visual imagery, and whether visual imagery merely constitutes the conceptual understanding of an object, or whether it is a genuine pictorial depiction of an object (i.e., propositional or depictive, respectively; Kosslyn et al., 2006). The main catalysts for this debate were due to the personal and internal nature of visual imagery, leading many to doubt its existence and suggest it was nothing but inner verbalisations of memory (i.e., propositional; Nanay, 2021; Thomas, 1997; Watson, 1928). However, arguments from Pearson and Kosslyn (2015) would bring the debate largely to a close in favour of imagery being a pictorial phenomenon: this, in light of the growing evidence from neuroscience that visual imagery formation, like visual perception, is aligned with activation in the primary visual cortex (for further details, see section *1.3*).

Visual imagery, and indeed mental imagery in general, can take a variety of forms. Differing from propositional attitudes, visual imagery adopts a depictive phenomenology, the content of which varies in terms of vividness and control. Visual imagery is understood to bear intentionality, insofar as the images possess some type of content. Relatedly, it can occur voluntarily and with purpose as well as occur as an involuntary and passive experience, relating visual imagery to subcategories of attention (endogenous and exogenous; Egeth & Yantis, 1997; Pearson, 2019). Such features are what make visual imagery resemble visual perception, thus coining it to be a "weak form of perception" (Pearson et al., 2015). However, despite concluding that visual imagery does indeed contain pictorial representations, the covert nature and ephemerality of visual imagery leads it to evade direct measurement in psychological research. Here, I argue that music listening, which is known to influence mental as well as other multisensory and physical experiences alike, may serve as a useful testing ground for expanding visual imagery research.

1.1.3. Music listening

Music listening is a ubiquitous human trait. Like language, music exists across all cultures and has developed to serve a multitude of purposes across time and cultures, such as emotional regulation, healing, social bonding, etc. (Cross, 2001; Stevens, 2012). In Western cultures, the development of modern sound technology has meant that music is an inevitable and omnipresent activity that one takes part in actively or passively (Juslin et al., 2008). Certain musical idioms appear to pique the natural interest of babies and infants, a capacity that gives rise to evolutionary arguments on the pre-historic role of music for communicating and relaying emotions (Jusczyk et al., 1993; Phillips-Silver & Trainor, 2005; Saffran & Griepentrog, 2001).

Beyond its role in human evolution, social bonding, and cultural practices, the richness of music and its absence of explicit semantics (unlike language) makes it an ideal stimulus with which to study complex cognitive phenomena and the functionalities of the human brain from several different disciplines. One particular strength of music is that it affords the embodiment of abstract concepts, presenting itself as a potentially rich source of information on the organisation of the brain when it is in the process of conjuring certain elusive mental phenomena. Indeed, there is a certain aboutness of non-lyrical music, that some posit is not grounded in concrete meaning and which leads to float between interpretations (Cross, 2011). In particular, imagination is thought to be the basis of musical perception and production, whereby music possesses a creative dimension that one derives aesthetic meaning from (i.e., musical imagination; Hargreaves, 2012). In the next section, I will introduce how music listening often leads to experiences of visual imagery, an imaginative process, in a listeners' mind's eye.

1.1.4. Visual imagery in music cognition research

Music is a rich stimulus with which to study the subjective and neural responses of visual imagery for a multitude of reasons. For instance, our ability to formulate metaphors and create connections between domains seems to be related to our experience of cross-modal associations to music (Deroy & Spence, 2016), a feature that can be readily related to imagery experiences to music (Deroy, 2020). The implications of this would shed light on our process of meaning-making to music (Leman et al., 2018) by investigating the types of imagery that might emerge as a function of that to music. Furthermore, music's acoustic features enable the investigation of imagery, an ephemeral process, against more discrete elements as they occur and unfold over time (Dahl et al., 2022; Herff et al., 2022; Juslin, 2019; Taruffi, 2021).

Research has shown that irrespective of musical genre (Markert & Küssner, 2021), mood, or listening situation (Juslin et al., 2008), visual imagery is a highly prevalent experience within experimental samples, with some researchers suggesting that over 70% of their samples reported experiencing it (Küssner & Eerola, 2019; Vuoskoski & Eerola, 2015). As a consequence of its prevalence, researchers have inquired into the content of experienced visual imagery. One content analysis showed that visual imagery appears to comprise landscapes, imagery of oneself, and memories, amongst other elements (Küssner & Eerola, 2019). Another in-depth investigation revealed it to additionally incorporate other more complex concepts, such as movement and luminescence (Dahl et al., 2022).

Indeed, when listening to music, individuals show the interesting ability to create narrative sequences out of the changing acoustic features in music. In their research on the effects of narrative descriptions of music on induced emotion, Vuoskoski and Eerola (2015) found that listeners are inclined to form visual imagery when provided with a narrative description of the music that they will be listening to. Visual imagery in response to music has further been suggested to be influenced by compositional features: Margulis (2017) noted that visual imagery responses varied significantly as a function of different composers' musical writing styles and musical genres. Additionally, Taruffi et al. (2017) found that listeners often imagined details referring to landscapes and natural elements when listening to sad music, whereas happy music evoked imagery related to dance, as well as motion in general.

Largely unknown to researchers however is whether the narratives that listeners form to music generally tend to converge across different listeners, or whether they are destined to be unique to the listener. In his influential report, Galton (1880) initially spoke of the automatic and habitual nature of visual imagery, represented by distinct and tractable characteristics, and even alludes to some degree of consistency amongst varied individuals:

> *"The conformity of replies from so many different sources, the fact of their apparent trustworthiness being on the whole much increased by cross-examination [...] and the evident effort made to give accurate answers, have convinced me that it is a much easier matter than I had anticipated to obtain trustworthy replies to psychological questions."* (p. 303).

Research has found most support for consistency in listeners' narratives. More recently, converging evidence from Margulis, Wong, et al. (2022) and McAuley, Wong, Bellaiche, et al. (2021) suggests that cultural background plays a significant role in the types and content of narrative visual imagery conjured, with a high level of consistency found amongst populations with shared cultural experiences. Such intersubjectivity demonstrates one example of the human capacity for core cognitive processes, such as scene reconstruction, future thinking, and episodic memory (D'Argembeau & Van der Linden, 2004, 2006; Hassabis et al., 2007; Rubin & Umanath, 2015). Additionally, the literature so far suggests that musical features such as contrast (sudden or unexpected musical changes; Margulis, 2017), rhythm and melody (Juslin, 2019), or even tempo (Herff et al., 2021) may play a part in influencing music-evoked visual imagery content. However, research is still limited in this regard.

Within music cognition, self-reports have been the most common tools for capturing visual imagery with only few studies, to date, having employed behavioural and physiological measures (Day & Thompson, 2019; Taruffi et al., 2017). As a result, there is a general lack of operational breadth in the study of the intrinsic and temporal features of music-evoked visual imagery (for a review of the advantages of studying the chronometry of music-evoked visual imagery, see Gelding et al., 2022). One broader goal of the current thesis is, therefore, to widen the scope of the methodological techniques used towards understanding music-evoked visual imagery.

1.2. LINKS BETWEEN MUSIC-EVOKED VISUAL IMAGERY AND AFFECTIVE STATES

Music listening is notably thought to encourage rich emotional associations. The affective experiences that individuals often feel when listening to music may not only impact the quality and outcome of the evoked visual imagery, but also emerge and develop from the content of visual imagery. Visual imagery possesses emotional connotations that result in an interplay between the visual, emotional, and auditory contents, the full-scale network of which is yet to be understood (Taruffi & Küssner, 2022). The usefulness of such a network has implications for situations such as performance, whereby musicians strategically conjure visual imagery to aid in portraying certain interpretations or sound qualities (Black, 2022; Presicce, 2022), or composition, whereby a film composer may wish to manipulate certain acoustic features in order to evoke powerful imagery or atmospheres (MacDonald, 2013).

Empirical research conducted on the topic of music-evoked visual imagery has largely focused on its relationship with emotion, with fewer investigations on this relationship occurring in the opposite direction. This section will provide a summary of this area of research and highlight current prevailing trends and outstanding questions.

1.2.1. Visual imagery and music-induced emotions

Most of the research on visual imagery to date has focused on identifying a relationship between it and music-induced emotions (e.g., Küssner & Eerola, 2019). The past couple of decades have seen an influx on visual imagery research in music cognition due to its inclusion in Juslin and Västfjäll's (2008) pivotal BRECVEMA framework of mechanisms underlying music-induced emotions (where other mechanisms include: *brain stem reflex*, *rhythmic entrainment*, *evaluative conditioning*, *contagion*, *episodic memory*, *musical expectancy*, and *aesthetic judgement*). One ecologically-valid study by Juslin et al. (2008) provides empirical support for visual imagery's involvement in emotion evocation through music. Adopting an experience sampling method for measuring everyday musical emotions, the authors showed that visual imagery was reported as the fourth (7%) most common cause of emotion induction, alongside episodic memory (14%), brain stem reflex (25%), and contagion (32%). This suggests that while visual imagery is prevalent in influencing emotions, it may not be a dominant mechanism in influencing music-induced emotions in everyday listening situations. This idea is further supported by Taruffi and Koelsch (2014), who found that visual imagery was a third most important mechanism in eliciting music-evoked sadness, after episodic memory, and emotional contagion.

Despite this, music-evoked visual imagery is often effective in enhancing emotions through the simulation of real events. The links between visual imagery and emotion induction through music have been demonstrated through the use of contextual information. In an early study by Stratton and Zalanowski (1992), they found that presenting listeners with classical music excerpts and instructions to imagine scenes with content that was mismatched from the excerpt composers' intended image led to less vivid visual imagery, compared to when asked to imagine images matching with the intended imagery of the music, and when in a silent condition with just images to imagine. Similarly, Vuoskoski and Eerola (2015) found that providing a sad contextual description alongside a sad-sounding musical excerpt led to enhancements in feelings of sadness, proposed to have occurred as a consequence of visual imagery potentially being conjured in response to the descriptions.

Recent neuroimaging research has provided further evidence that visual imagery acts as a representational mode for other common forms of spontaneous cognition induced in response to music varying in valence (Martarelli et al., 2016). In a study by Taruffi et al. (2017), participants underwent functional magnetic resonance imaging (fMRI) and listened to instrumental music evoking feelings of happiness and sadness. Participants reported on the content of their mind wandering, including the form in which it took place. They found that the excerpt evoking sadness increased the strength of mind wandering, and was associated with greater centrality in the central node of the Default Mode Network (DMN), a key contributor to mind wandering (Andrews-Hanna, Reidler, Huang, et al., 2010; Andrews-Hanna, Reidler, Sepulcre, et al., 2010; Christoff et al., 2009). Importantly, however, visual imagery was the predominant modality in which the mind wandering took place (as opposed to inner speech) in response to both happy and sad music. Moreover, in an EEG study by Fachner and colleagues (2019), observations of the temporal dynamics of visual imagery- and emotion-related indices suggested that visual imagery that occurred during moments of interest coincided with emotions that varied in valence and intensity, as reflected by variations in frontal alpha asymmetry.

Though there is strong empirical support for the link between visual imagery and music-induced emotions, directionality of this relationship is still largely unknown. Mechanisms with inherently more measurable traits lend themselves to be identified more easily through the manipulation of certain musical events. For instance, Juslin et al. (2014, 2015) were able to activate four of the BRECVEMA mechanisms, brain stem reflex, emotional contagion, episodic memory, and musical expectancy, by altering parameters in the music than these mechanism provide psychological meaning for (e.g., activating musical expectancy by altering a piece of music to violate melodic and harmonic expectations). For causal relationships between visual imagery and music-induced emotions to be identified, more objective and implicit empirical paradigms should be sought after and embraced.

1.2.2. Causal relationships between visual imagery and affective states

One outstanding question that persists amongst research relates to the directionality of the relationship between visual imagery and music-induced emotions. In most cases, assessment of these experiences in research takes place simultaneously, meaning that researchers are unable to conclude on whether visual imagery tends to occur before or after emotion induction. Even so, a handful of studies have begun to provide insight on their temporal relationship by adopting indirect approaches that assess their occurrence over time. For instance, Day and Thompson (2019) investigated the temporal relationship between visual imagery and music-induced emotions using response times as an implicit measurement. Participants listened to short classical or pop music excerpts and were asked to indicate via a key press as soon as they (1) recognised the emotions expressed by the music, (2) experienced an emotional response to the music, and (3) experienced visual imagery in response to the music. They showed that perceiving an emotion resulted in the shortest response time, followed by feeling an emotion, and then experiencing visual imagery. Despite these findings, Taruffi and Küssner (2019) propose that these results could be due to the fact that their musical stimuli conveyed low-level affect whereby processing is less cognitively taxing than more complex emotions such as nostalgia. Thus, the time taken to experience emotions and visual imagery could be largely governed by the complexity of the emotions or the contents of the imagery.

This sentiment was somewhat addressed in research by Vreogh (2018, 2021), who suggests an alternative viewpoint whereby the *quality* of the emotion is what mediates the relationship between visual imagery and emotion induction during music listening. In a large-scale online survey, Vroegh (2021) estimated network models in order to understand the interdependencies of different states of consciousness, felt emotions, and visual imagery. Participants listened to a piece of music from a genre of their own choosing, and reported on the quality of their felt emotion, phenomenological consciousness, and visual imagery prevalence and vividness. His analysis revealed that positive affect, internally-directed attention, and altered experience were central in predicting visual imagery, which, in turn, was found to influence short- and long-term memory and mixed affect.

In all, these studies demonstrate that an elusive experience such as visual imagery may require the implementation of sophisticated measurements and analytical techniques if one is to understand the conditions that lead to its occurrence, and how qualities of its content elicit further cognitive phenomena.

1.3. NEURAL BASES OF MUSIC-EVOKED VISUAL IMAGERY

It is widely recognised that part of the difficulties with investigating visual imagery lies in its private nature. Indeed, self-report measures of visual imagery, while incredibly useful for uncovering the qualitative elements of imagination, are arguably heavily relied on. Research findings that individuals are indeed aware of their own meta-cognitive states (Pearson et al., 2011) and can accurately validate their own subjective reports (Seli et al., 2015) are growing and important. However, neuroscientific measures are central, and remain necessary to provide tangible observable evidence of the experience of visual imagery, including in response to perceptual stimuli such as music.

1.3.1. Visual imagery versus visual perception

There have been significant advances in our understanding of the neural underpinnings of generally-occurring visual imagery in the recent decade (Borst & Kosslyn, 2008; Dijkstra et al., 2019; Pearson, 2019; Sousa et al., 2017; Xie et al., 2020). It is now well documented that visual imagery involves similar connectivity patterns to those implicated in visual perception (Cichy et al., 2012; Dijkstra et al., 2019; S.-H. Lee et al., 2012; Mutha et al., 2014; Xie et al., 2020; Zacks, 2008). Specifically, research has found similarities between imagery and perception in terms of activity in the primary visual cortex, especially with regard to content-related information governed by top-down mechanisms (Dijkstra et al., 2019; Ishai et al., 2000).

However, research increasingly suggests that there are nuanced differences in the activation of early visual areas during the perception and imagery of low-level features. Specifically, it is suggested that perception and imagery activate bottom-up and top-down processes differently depending on whether there is a need to engage cognitive resources (i.e., retrieval of sensorial representations in perceived information vs. retrieval of information from long-term memory in imagery; Mechelli et al., 2004). Most recently, a review by Dijkstra (2024) has suggested that the early visual cortex is only selectively activated during imagery when there is a required level of detail needed to imagine a given object. Studying the neural correlates of visual imagery during music listening offers a new and valuable opportunity to examine the extent to which visual imagery recruits lower-level perceptual processing areas.

1.3.2. Implicated brain areas and frequency bands during visual imagery

Visual imagery is thought to be processed across a network of brain areas necessary for the aid of mental rotation and object memory recall (for a review, see Pearson, 2019). Early to more modern studies have tended to emphasise the role of alpha oscillations [8-13 Hz], responsible for reflecting internally directed (Salenius et al., 1995; Williamson et al., 1997; Xie et al., 2020), within the posterior brain regions where the visual cortex is located (Drever, 1955; Gale et al., 1972; Kaufman et al., 1990; Williamson et al., 1997) in visual imagery. Generally speaking, as for visual perception, visual imagery of complex content may implicate a wide range of oscillatory frequency bands and brain areas.

With regard to implicated areas, a large body of studies has shown that (in addition to posterior areas), parietal, central, and frontal areas of the brain are also involved in visual imagery, particularly in the context of spatial and motor features (de Borst et al., 2012; Menicucci et al., 2020; Mutha et al., 2014; Sousa et al., 2017; Thompson et al., 2009; Villena-González et al., 2018; Zabielska-Mendyk et al., 2018; Zacks, 2008). For instance, mental rotation tasks, which often involve the visual imagination of objects in movement, have been shown to be associated with increased activity in parietal and precentral areas that are involved with motor simulation (Thompson et al., 2009; Zacks, 2008). Further, it has been suggested that frontal regions may be essential for integrating the “what” and “where” contents of visual thought, that are processed by occipitotemporal and parietal areas respectively (de Borst et al., 2012). Finally, in a meta-analysis seeking to clarify what, if any, shared components exist for kinaesthetic and visual imagery in the context of sports (Filgueiras et al., 2018), it was shown that athletes’ visual motor imagery (i.e., visualising a movement execution) of sports actions was similar to their kinaesthetic motor imagery (i.e., imagining the sensations of the movement execution) in recruiting, amongst others, frontal motor (including premotor and supplementary motor areas) and parietal areas involved in feeling their own movements (somatosensory cortex, inferior and superior parietal lobule). The authors argued that the surprising finding of somatosensory cortex activity being present during visual motor imagery may be because athletes cannot help but have their visual imagery influenced by their bodily sensations.

Furthermore, in addition to the EEG studies emphasising *suppression* of occipital alpha (indicating increased neural firing in visual areas) as a signature of visual imagery formation, enhanced gamma power [30-45 Hz] in the occipital brain region has also been associated with the experience of creative and vivid spontaneous visual imagery (Luft et al., 2019), the content-specific features of visual imagery (Lehmann et al., 2001), and with working memory load during motor imagery (De Lange et al., 2008; Sepúlveda et al., 2014). Similarly, theta [4-7 Hz] and beta [14-30 Hz] oscillations have (along with alpha) been shown to be successful in discriminating the contents of visual imagination (Xie et al., 2020), while lower [8-10 Hz] and upper [11-13 Hz] bands of alpha appear to show nuances in terms of how they relate to visual imagery formation (Gualberto Cremades, 2002; Petsche, 1996).

1.3.3. The neural correlates of music-evoked visual imagery

Our knowledge on whether music-evoked visual imagery elicits similar neural responses to imagery in response to other perceptual stimuli is currently highly limited (for a general overview of neuroscientific measures of music and mental imagery, see Belfi, 2022). Despite the relevance of studying the neural correlates of the content of music-evoked visual imagery, only one such study exists: Fachner et al. (2019) recorded dual-EEG during a guided imagery and music (GIM) therapy session – where GIM involves the induction of visual imagery in response to a specialised GIM soundtrack – from both a therapist and client simultaneously, with the purpose of observing a potential coupling of visual imagery – and emotion-related processes. They identified and compared moments of interest (characterised by client-ascribed “pivotal moments” of imagery formation) to moments of non-interest, they found greater posterior alpha suppression during moments of visual imagery formation. EEG thus presents itself as a reliable measurement technique for detecting occurrences of visual imagery.

1.4. INDIVIDUAL DIFFERENCES THAT INFLUENCE VISUAL IMAGERY

Furthermore, in line with their contributions in a range of other domains, there is a need to consider the implications of inter-individual factors that may shape the experience of music-evoked visual imagery. Even variations in the imagined surface features of an object, such as vividness, are said to influence the extent to which activation of the early visual cortex is found (Cui et al., 2007; Dijkstra, Bosch, et al., 2017). Thus, in addition to identifying objective and indirect markers of music-evoked visual imagery to emphasise subjective reports, it is also important to understand the intrinsic attributes of a listener that determine the strength and quality of their imagery.

For instance, literature has found evidence that individuals with a higher disposition for empathy engage in higher levels of visual imagery in response to music (Balteș & Miu, 2014; Taruffi et al., 2018, 2021). Furthermore, there is some indication that music possesses social implications for forming strong emotional images and consolidating one’s personal identity (Denora, 2010; Juslin, 2019; Larson, 1995); for instance, those in late adolescence use music to create and foster external “trendy” impressions or “images” of themselves (North et al., 2004). However, in this thesis, I will refer specifically to the individual differences that have either received minimal attention in literature (but are continually growing in breadth) as well as others that have yielded inconsistent results.

1.4.1. Influence of musical training

Aside from the established evidence that musicians make use of motor or kinaesthetic imagery in preparation for performances (Bernardi et al., 2013; Keller, 2012; Lotze & Halsband, 2006; Schuster et al., 2011), and the constructive role it has in the practice and participation in musical performance by enabling anticipation and action planning (Floridou, 2022), there is evidence that visual imagery poses additional emotional implications on musicians compared to non-musicians. For example, Küssner and Eerola (2019) showed that music-evoked visual imagery was characterised by two clusters: vivid and soothing visual imagery. They suggest that one of the main reasons that musicians use visual imagery was to modulate their arousal levels through, potentially, increasing it using vivid visual imagery or decreasing it using soothing imagery. However, further research is needed to support visual imagery's influence in musicians.

1.4.2. Aphantasia: a special population of non-visualisers

Many individuals use visual imagery to make sense of their surroundings. For instance, they create mental schemas of the environment to consolidate memories of past events and situations (Holmes et al., 2008). However, for a small percentage of the general population (2.1 – 2.7%, Faw, 2009; 2 – 5%, Galton, 1880), attempts to visualise an object lead to no sensorial experience of it. The term to denote this phenomenon, *aphantasia*, was originally coined by Zeman et al. (2015), who described the condition as being characterised by "reduced or absent voluntary imagery" (p. 4). Some individuals do not ever recall being able to form images in their mind, which implicates the idea that they may possess a congenital form of aphantasia (amongst other types: neurogenic, psychogenic; Zeman, 2020), which some research has associated with impaired performance in other imagery tasks (Farah, 1984); whereas others develop "acquired" aphantasia: for instance, Zeman et al. (2010) describes the case study of "MX" who experienced a sudden loss of his visual imagery ability. fMRI scanning revealed, during the attempted imagery of famous faces, significantly reduced activation in a network of posterior brain regions, but increased activity in frontal regions, potentially reflecting the newly learned use of alternative strategies to perform visuospatial tasks.

Research into aphantasia so far has tended to rely on self-report questionnaires such as the Vividness of Visual Imagery Questionnaire (Marks, 1973) or the Plymouth Sensory Imagery Questionnaire (Andrade et al., 2014) as diagnostic tools. Based on this need to rely on self-report and subjective accounts from aphantasics, it has been debated whether aphantasia is characterised by a real absence of visual imagery or by poor metacognitive and introspective abilities (Keogh, Pearson, et al., 2021; Keogh & Pearson, 2011). Recently, however, objective evidence of poor visual imagery ability has been suggested by neuroscientific and physiological data. For instance, when asked to imagine famous faces and buildings in an fMRI study, low visual imagers activated a widespread set of brain areas negatively associated with imagery vividness when compared to high visual imagers, who predominantly showed restricted brain activations in the posterior cortices (Fulford et al., 2018). Similarly, Kay et al. (2022) asked aphantasics and controls to compare the brightness of a grey square to a prime, and found that aphantasics did not display the typical pupillary light response (a reflex of the pupil to expand and constrict in response to changes in light) in response to illusory brightness that the control group did.

In the context of music, the inability to form visual imagery poses potential implications on the strength of other cognitive experiences, such as emotional evocation (Blackwell, 2020). In turn, there is a potential that having aphantasia necessitates alternative uses of music or a predominance of other mechanisms in place of visual imagery for forming emotional connections. This question is addressed in Studies 5a and 5b (Chapter 6) of this thesis.

1.5. PRACTICAL USES OF MUSIC-EVOKED VISUAL IMAGERY

Like visual imagery, other conscious visual experiences such as synaesthesia and hallucinations, also manifest across a spectrum of intentionality. Visual imagery is one phenomenon with the potential to occur both voluntarily (i.e., deliberately) or involuntarily (i.e., spontaneously). Voluntary and conscious imagery is often applied as a tool in learning processes in various situations (e.g., in the case of motor imagery during music pedagogy to support movement rehabilitation, Schaefer, 2022). However, the extent to which visual imagery is conjured by individuals as an involuntary cognitive strategy with potential to alleviate anxious or stressful states has rarely been addressed in the literature. Indeed, a range of psychological and neurological disorders are underlied by (voluntary and involuntary) visual imagery. Their rehabilitation would greatly benefit from an enhanced empirical understanding of the effects of the control of visual imagery, thus divulging new treatment avenues. Music, with its ability to evoke, enhance, and influence visual imagery content, possesses potential for forwarding such aims. The current section briefly compares visual imagery states varying in intentionality, and then discusses the implications of utilising music-evoked visual imagery in therapeutic settings.

1.5.1. Voluntary (deliberate) and involuntary (spontaneous) visual imagery

In order to understand the feasibility of applying visual imagery within therapeutic contexts, one important factor to consider is the differences and utility of different subtypes of visual imagery pertaining to intentionality, i.e., spontaneous and deliberate. Each subtype possesses its own positive and negative implications on an individual (Pearson, 2020). In one sense, encouraging voluntary visual imagery has clinical implications for dealing with intrusive memories and past traumas by bringing "images to consciousness" (Taruffi & Küssner, 2019). In the other, involuntary visual imagery is a core symptom of certain mental disorders (such as post-traumatic stress disorder; PTSD). The underlying neural circuitry that influences each subtype of visual imagery is yet to be understood, both within the music cognition literature and in neuropsychology more generally. However, a framework by Pearson and Westbrook (2015) characterise voluntary and involuntary imagery manifestations in terms of top-down and bottom-up processing, respectively.

Similar to the general difficulties in measuring the temporal fluctuations in visual imagery, fluctuations in episodes of visual imagery intentionality are doubly unpredictable. Studies 3 and 4 (Chapters 4 – 5) address this issue by proposing an alternative methodological approach borrowed from literature on mind wandering (Polychroni et al., 2022; Weinstein, 2018).

1.5.2. A mechanism in alleviating negative emotional states

Visual imagery plays an influential role in determining various psychological states; Holmes et al. (2008) emphasised the crucial role it plays in clinical settings by suggesting that it works as an "emotional amplifier" in cases of anxiety. On the one hand, visual imagery is cited as a hindrance and a threat to one’s mental wellbeing; for instance, in the mental replaying of vivid intrusive images as a result of PTSD. In such cases, research has found that limiting or suppressing visual imagery formation can lead to a decrease in the number of episodes of intrusive imagery; this, by way of administering tasks with the capability of cognitively overloading the visuo-spatial sketch pad and thus impeding one’s inclination to creating intrusive thoughts (Andrade et al., 1997; Holmes et al., 2004; Kavanagh et al., 2001; Kemps & Tiggemann, 2007). Such tasks have similarly been shown to be effective in diminishing visual imagery experienced in response to music; a study by Hashim et al. (2020) demonstrated distinct decreases in visual imagery prevalence and vividness, as well as mild attenuations in emotional response ratings, following a lateral eye-movement distractor task in comparison to a no-distractor control task.

The application of visual imagery within therapeutic settings has also been shown to possess benefits through communicating the power of music. More specifically, some research propounds its potential ability to promote feelings of social bonding. In a study by Herff et al. (2023), participants performed a directed mental-imagery task, whereby they imagined a figure on a journey towards a topographical landmark either in silence or while listening to task-irrelevant music. Using computational modelling approaches, the authors were able to identify latent topic structures differentiating content described in response to the music or in response to silence and found that one prominent topic that arose pertained to social interaction.

Additional EEG research has emphasised that there is a basis to visual imagery’s ability to facilitate improved wellbeing. Music’s ability to encourage personal and rich visual imagery is advantageous for imagery-based therapies, such as Bonny’s Guided Imagery and Music (GIM) therapy, that seeks to extract common themes and underlying emotional issues from guided visual imagery formed by a patient (Bonny, 1983, 1986; Fachner et al., 2019; Fox & McKinney, 2016; Grocke, 2010). The benefits of such content have been made evident from objective measurements such as EEG. As mentioned earlier, Fachner et al. (2019) followed the GIM sessions of two patients and recorded the brain activity of the interaction between the therapist and patient. They identified moments of interest in the session characterised by target sequences regarded as relevant and important, and analysed brain frequency signals pertaining to visual and emotion processing. They found significant alpha power suppression during moments of interest to signify the presence of visual imagery. By tapping into patients' visual imagery content during these sessions they demonstrated themes related to loss and closure, suggesting that visual imagery during music listening can indeed play a key role in wellbeing.

1.6. SUMMARY, RESEARCH QUESTIONS, AND THESIS OUTLINE

To summarise, I have outlined that research into music-evoked visual imagery has presented rich findings on its relationship with affective responses to music (Balteș & Miu, 2014; Belfi, 2019; Day & Thompson, 2019; Hashim et al., 2020; Vuoskoski & Eerola, 2015), with increasingly promising results on the inherent characteristics of music-evoked visual imagery (Dahl et al., 2022; Herff et al., 2021, 2022; Küssner & Eerola, 2019), as well as its potential uses in therapeutic contexts (Fachner et al., 2019; Herff et al., 2022, 2023; Panteleeva et al., 2018). While this is an important factor especially in how we experience music in our day-to-day lives, as demonstrated above, visual imagery has the potential to offer implications into alleviating feelings of stress by acting as a mechanism through which music is relaxing. In addition, there is a notable lack in our understanding of the brain's responses to visual imagery when formed during music listening, and research in this regard would aid in establishing objective techniques identifying visual imagery formation to music in the first instance, and subsequently understanding its potential interactions with other cross-modal domains and physiological states.

Throughout this thesis, I will present a comprehensive series of work that draws upon an array of methodological approaches to investigate current gaps within the research of music-evoked visual imagery. Across five empirical investigations, this thesis aims to address four overarching research questions:

1) What are individuals visualising when they listen to music, and how consistent is the visual imagery within themselves and with others? [Chapter 2]

2) How are different forms of music-evoked visual imagery represented in the brain? [Chapters 3 – 5]

3) Does visual imagery provide any emotional benefits or serve a relaxation function when listening to music? [Chapter 5]

4) To what extent do aphantasic listeners' emotional and aesthetic experience of music differ from that of a general population, as a result of an inability to experience visual imagery? [Chapter 6]

The first study of this thesis within Chapter 2 will present a detailed thematic investigation into what we visualise during music listening and the consistency of such imagery when compared to ourselves and other individuals. Study 2 in Chapter 3 offers a neuroscientific outlook by examining how static and dynamic types of music-evoked visual imagery are reflected in EEG data, with particular focus on replicating past effects of activity in the visual cortex, as well as potential activity in motor cortices reflecting action-based imagery. Study 3 in Chapter 4 builds on this by further using EEG to identify neuro-oscillatory clusters representing spontaneous and deliberate forms of visual imagery during extended listening tracks that vary in terms of relaxation potential and familiarity. Next, Study 4 in Chapter 5 utilised EEG once more, as well as physiological skin conductance response and self-report methods, to discuss music-evoked visual imagery's potential role in alleviating feelings of stress. In a final two-part investigation, Studies 5a and 5b in Chapter 6 revisited visual imagery's purported role as a mechanism influencing music-induced emotions by comparing the affective responses and everyday uses of music between individuals with *aphantasia* and individuals from the general population, with the aim of providing a rudimentary understanding of the extent of visual imagery's influence in affecting other hedonic response to music. This thesis will end with Chapter 7, which will summarise all the key and novel findings of this research. This final chapter will go on to critically outline key limitations of the presented research, before going on to discuss important implications on the wider research of imagination, visual imagery, and music listening more specifically, and will end with the consideration of a few future extensions to the presented works.

CHAPTER 2. INVESTIGATING THE CONTENT AND CONSISTENCY OF MUSIC-EVOKED VISUAL IMAGERY

To date, there have been few thorough investigations into the specific content of music-evoked visual imagery, and whether listeners exhibit consistency within themselves and with one another regarding their visual imagery content. In Study 1, I recruited an online sample (N = 353) who listened to three orchestral film music excerpts representing happy, tender, and fearful emotions. For each excerpt, listeners rated how much visual imagery they were experiencing and how vivid it was, their liking of and felt emotional intensity in response to the excerpt, and, finally, described the content of any visual imagery they may have been experiencing. Further, they completed items assessing a number of individual differences including musical training and general visual imagery ability. Of the initial sample, 254 respondents completed the survey again three weeks later. A thematic analysis of the content descriptions revealed three higher-order themes of prominent visual imagery experiences: Storytelling (imagined locations, characters, actions, etc.), Associations (emotional experiences, abstract thoughts, and memories), and References (origins of the visual imagery, e.g., film and TV). Although listeners demonstrated relatively low visual imagery consistency with each other, levels were higher when considering visual imagery content within individuals across timepoints. The findings corroborate past literature regarding music's capacity to encourage narrative engagement. It, however, extends it (a) to show that such engagement is highly visual and contains other types of imagery to a lesser extent, (b) to indicate the idiosyncratic tendencies of listeners' imagery consistency, and (c) to reveal key factors influencing consistency levels (e.g., vividness of visual imagery and emotional intensity ratings in response to music). Further implications are discussed in relation to visual imagery's purported involvement in music-induced emotions and aesthetic appeal.

2.1. INTRODUCTION

Listening to programmatic excerpts of music, such as *Peter and the Wolf* by Sergei Prokofiev, may elicit the imagination of several different scenes. For some listeners, it might enliven childhood memories of first encounters with the piece, while for others, it may render images of the wolf, dark forests, and feelings of threat. In some cases, particular themes represented by different instruments may even give rise to the visualisation of specific characters from the tale (Maus, 1991; Newcomb, 1987). All are plausible situations afforded by listening to the piece. However, such a plurality of possibilities raises the question of the extent to which listeners are prone to the same mental experiences when listening to a piece, or whether listeners' life-long learned experiences are too complex to allow for such shared visual imagery.

For most listeners, forming a narrative in their mind's eye is a way to engage with heard music (Küssner & Eerola, 2019; Margulis, 2017), and such narrative sequences are often reported to be vivid and multi-thematic experiences (Herff et al., 2021; Margulis et al., 2022). A few investigations into visual imagery content during music listening have begun to shed light on this idea of imagery consistency across listeners (Dahl et al., 2022) and potential influencing factors (Margulis, Wong, et al., 2022). However, the extent to which listeners exhibit similarities in their own visual imagery across listening situations is still an open question.

2.1.1. Semantic associations and what listeners imagine

Musicologists commonly speak of a "narrative dimension" (Nattiez, 1990) to music and refer to the explicitness through which this might be communicated by drawing parallels between music and language (Levinson, 2004; Maus, 1991; McClary, 1993; Nattiez, 1990). Research has shown that semantic associations inform our understanding and perception of music, and that the process of narrativizing music is rooted in, and requires access to, linguistic cognitive resources. A study by Koelsch et al. (2004) revealed that priming words with musical stimuli induced the same electrophysiological signature (the N400) as when priming with linguistic stimuli, suggesting that as with language, listeners are capable of extracting meaning from musical stimuli. Interestingly, though, despite links between music and language processing being repeatedly demonstrated (Cross, 2011; Raffman, 1993; Steinbeis & Koelsch, 2008b; Sternin et al., 2021; Tillmann, 2012), they seem to nevertheless also be subserved by distinct cognitive mechanisms: Barraza et al. (2016) used EEG to reveal that while linguistic processing appears to work on the basis of confirming participants' prior expectations for word primes, music listening implicates a post-hoc cognitive strategy, focused more on forming meaning where, due to the often subjective nature of deriving musical meaning, there is a *lack* of it. Such findings suggest that visual imagery in response to music may be a loose and unstructured process.

Given the fact that semantic associations can be formed from music, an interesting question that follows is the type of content that listeners imagine in response to music. In an online survey, Küssner and Eerola (2019) asked participants to provide descriptions of the visual imagery that they typically experience while listening to music. They found that the most common types of visual imagery were natural landscapes and personal memories. Additional content included images of musical performances as well as abstract types of visual imagery such as colours and geometric shapes.

Empirical research demonstrates that people's thoughts when listening to music are generally often visual in nature, with mind wandering, or 'daydreams', tending to occur more often in the form of visual images than in the form of words (Koelsch et al., 2019; Taruffi et al., 2017). While examining the effects of music on thought content and valence, Koelsch et al. (2019) found visual imagery to be a characteristic of mind wandering in response to music that evokes heroism and sadness. An investigation into the metaphors evoked by music carried out by Schaerlaeken et al. (2019) revealed that these could be represented by five main attributes (Flow, Movement, Force, Interior, and Wandering) with visual imagery emerging as a prominent mode in which they are realised. Studying visual imagery in response to music promises theoretical insights into how cross-modal associations (Deroy & Spence, 2016) and meaning making emerge (Leman et al., 2018). Taken together, these studies provide some clues into the types of visual imagery content that may be expected to be derived from music listening.

2.1.2. How stable is music-evoked visual imagery?

To date, visual imagery has been assumed to be a largely idiosyncratic experience, with listeners prone to conjuring mental images not only through acoustic features, but also through processes involving past life events and contextual information (Juslin, 2019). Yet, the notion of consistency has been sparingly addressed in music-evoked visual imagery research.

Recent related work by Margulis et al. (2022) has investigated the formation of imagined narratives – mental storytelling being formed with the potential (but not necessity) to appear visually – and suggests that imagined narratives may be more widely shared than previously thought. With two sets of independent samples recruited from the USA and China, they used natural language processing techniques to analyse open-text reports of narratives formed in response to instrumental music. They found that descriptions written by individuals who shared an underlying culture received higher consistency levels than reports written by those who did not.

These results offer strong support for the semantic similarities that individuals can have in their imagined narratives to music. However, it is important to note that consistency there was made in relation to general narrative formation and not with respect to visual imagery. Thus, it is possible that focusing specifically on music-evoked visual imagery would yield a less pronounced similarity level.

In an influential paper, Cross (2011) proposes that music possesses a *floating intentionality* whereby different listeners derive different meanings from the ongoing events in music. In line with the work of Margulis and colleagues, one might expect that individuals will tend to show higher consistency of visual imagery within themselves than with one another. However, such a possibility has never been empirically tested. One aim of the current study was therefore to shed light on this matter by comparing the degree of similarity in listeners' visual imagery reports within themselves (across two surveys) to the degree of similarity in visual imagery reports they show with other listeners.

2.1.3. Emotion and aesthetic appeal

Previous literature hints at the idea that one's experience of visual imagery may somewhat be related to their hedonic responses to music. Music-evoked visual imagery has been previously associated with aesthetic and emotional engagement with music, but the evidence of such links remains limited (Juslin, 2013; Taruffi & Küssner, 2019). In one study, Belfi (2019) investigated factors contributing to the aesthetic appeal of classical, jazz, and electronic music. Participants were required to listen to tracks taken from each genre and to report on their experience with respect to the vividness of music-evoked visual imagery, arousal, emotional valence, and liking/aesthetic appeal of the music. Across genres, it was found that emotional valence and visual imagery vividness were similar in the degree of their predictive influence over musical aesthetic appeal.

In another study, Presicce and Bailes (2019) established a clear association between the continuous ratings of visual imagery that participants reported in response to a selection of piano pieces and their continuous ratings of engagement with the music. Finally, in an fMRI study, Koelsch et al. (2013) showed evidence of an interaction between emotion and visual areas of the brain during music listening. Interestingly, listening to fearful, compared to joyful music, led to greater interactions between the visual cortex and the superficial amygdala, perhaps due to the heightened vigilance fearful music elicits. Thus, associations between music-evoked visual imagery and emotion induction are evident, although the directionality of this relationship is still a widely debated topic supported by contrasting (Day & Thompson, 2019; Vroegh, 2018).

As previous studies have stated that extensive prior experience with the visual arts may be expected to influence the experiences of visual imagery that individuals have (mental imagery being regarded as a useful tool in creative processes including the visual arts, music, and dance; see Rosenberg & Trusheim, 1989; Wong & Lim, 2017), the current study also investigated how the prevalence, vividness, and consistency of music-evoked imagery reported is influenced by this individual difference.

Taken together, it is apparent that visual imagery may be at least one tool with which one is ‘drawn in’ to music and by which music engages and induces emotions in a listener. However, there is a need to explore this with a larger sample than has been used in previous research and with a more nuanced approach that can discern particular types and subtypes of visual imagery during music listening.

2.1.4. The present study

In light of the reviewed literature, the present work sought to address four main aims:

1) To run a thematic analysis to create a hierarchical framework highlighting the prevalent codes and overarching themes found in descriptions of music-evoked visual imagery content,
2) To ascertain the extent of the consistency of visual imagery within and across individuals during music listening,
3) To examine potential behavioural factors and individual differences (general visual imagery ability, musical training, and participation in the visual arts) that may be driving the rates of within- and across-participant consistency levels,
4) To test the extent to which the prevalence and vividness of visual imagery is associated with the emotional intensity and aesthetic appeal of music, as well as an array of individual differences (general visual imagery ability, musical training, and participation in the visual arts).

Importantly, it was anticipated that storytelling, including action-based imagery (Schaerlaeken et al., 2019, but see also Eitan & Granot, 2006; Johnson & Larson, 2003), may emerge as a prominent form of visual imagery (H1) (Küssner & Eerola, 2019; Margulis, 2017; Margulis, Wong, et al., 2022), and that while, in line with Cross (2011), within- and across-participant consistency would be shown to be modest, there would be higher within-participant consistency than across-participant consistency (H2), due to the reduced set of factors that can lead to variance (i.e., no differences in individual listening experience or personal memories when comparing individuals with themselves, in contrast to when comparing with different individuals).

Further, it was predicted that previous reports of a relationship between visual imagery and both emotion induction (Day & Thompson, 2019; Hashim et al., 2020; Juslin & Västfjäll, 2008; Vuoskoski & Eerola, 2015) and aesthetic appeal (Belfi, 2019) would be replicated. It was specifically predicted that the prevalence and vividness of visual imagery would both be predicted by ratings of emotional intensity (H3) and music liking (H4), in line with previous reports of positive links between these phenomena (Belfi, 2019; Hashim et al., 2020; Juslin, 2013, 2019).

Next, with regard to inter-individual differences, it was predicted that the prevalence and vividness of visual imagery in response to music would be strongly associated with general visual imagery abilities (H5) but show a weaker relationship with musical training (H6); this is due to previous findings that musical expertise had little predictive influence over visual imagery content (Herff et al., 2021) and that while musically trained individuals may possess higher imagery abilities, this is not necessarily true of *visual* imagery (Talamini et al., 2022).

Finally, it was nevertheless predicted that the higher imaginative freedom of those who regularly engage with the visual arts may allow them to experience music-evoked visual imagery more frequently and vividly than those who do not regularly engage with the visual arts (H7).

2.2. METHODS

2.2.1. Participants

353 participants (153 female, 198 male, 2 prefer not to say) aged 18-66 years (M = 26.41, SD = 9.41) were recruited using either Prolific, an online participant recruitment platform, or word-of-mouth. The same sample was invited three weeks later to once more take part, with 254 participants (102 female, 149 male, 3 prefer not to say) aged 18-66 (M = 26.85, SD = 9.04) retaking the survey. See Table 2-1 for further demographic information.

Table 2-1

Countries of residence of the samples from survey 1 and survey 2.

Country of Residence	**Survey 1 (N = 353)**		**Survey 2 (N = 254)**	
	Sum	**Proportion**	**Sum**	**Proportion**
United Kingdom	79	22.4%	38	15.0%
Portugal	59	16.7%	48	18.9%
Poland	48	13.6%	37	14.6%
Italy	18	5.1%	16	6.3%
Canada	18	5.1%	14	5.5%
Mexico	17	4.5%	15	5.9%
Germany	14	4.0%	3	1.2%
Greece	13	3.7%	12	4.7%
Other	87	24.9%	71	28.0%

2.2.2. Ethics Statement

This research has received ethical approval from the Ethics Committee of the Department of Psychology at Goldsmiths, University of London. Participants provided online written consent and received monetary compensation or course credit for their time.

2.2.3. Materials and stimuli

Three film music stimuli conveying happy, tender, and fearful emotions were selected from Eerola and Vuoskoski's (2011) database (see, https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/emotion/soundtracks for access to the original stimuli). These excerpts were obtained from the catalogue of extended 1-min film excerpts (see Appendix of Vuoskoski et al., 2012), or see, https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/emotion/soundtracks-1min for access to the original stimuli), validated to be unfamiliar to most listeners and to still convey the intended emotions even in their shorter form. In terms of the films that the tracks were taken from, the excerpt conveying happy emotions was taken from *The Untouchables* soundtrack (track 6, number 071 from Eerola and Vuoskoski's set of 110 tracks). The tender excerpt is from the *Shine* soundtrack (track 10, number 042 from set of 110 tracks). Finally, the fearful excerpt is from the *Batman Returns* soundtrack (track 5, number 011 from set of 110 tracks). In order to ensure uniformity amongst the musical excerpts, as well as to control the overall length of the survey, all excerpts were edited to last a duration of 45 seconds using Audacity (Version 2.3.2.0). These were also edited to finish with a fade-out to avoid an abrupt ending.

Visual imagery ratings were obtained using two items from Pekala's Phenomenology of Consciousness Inventory (Pekala, 1991), a 53-item questionnaire that measures a variety of personal perceptual experiences revolving around consciousness. Participants were asked about the prevalence of visual imagery in their experience (1 = I experienced no visual imagery at all, to 7 = I experienced a great deal of visual imagery), and the vividness of their imagery (1 = My visual imagery was so vague and diffuse, it was hard to get an image of anything, to 7 = My visual imagery was so vivid and three-dimensional, it seemed real).

The content of visual imagery was measured using an open-text question. Participants were asked to describe the content of their visual imagery (if any) with no limit to the length of their descriptions required. Liking ratings in response to the music were measured using a 5-point Likert scale, from 1 = Dislike a great deal, to 5 = Like a great deal. Emotional intensity ratings were also measured using a 5-point Likert scale, from 1 = Not at all intense to 5 = Extremely intense. Participants were also asked to report on the musical aspects they thought contributed to their visual imagery, but data from this question will be addressed elsewhere.

Finally, three aspects of individual differences were measured. The Musical Training dimension of the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) was administered in order to gauge the extent of individuals' musical experience. The Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973), which comprises 16 statements to which individuals are instructed to form a visual mental image in their minds, was administered as an independent measurement of visual imagery ability, along a 5-point Likert scale from 1 = No image at all, you only 'know' that you are thinking of an object to 5 = Perfectly clear and vivid as real seeing. Finally, any experience with activities associated with the visual arts ('*Do you participate in any activities associated with the visual arts?*') was also probed. This was a binary yes/no question, with an option to elaborate on the type of activity experienced if answered in the affirmative.

2.2.4. Procedure

Participants were first provided with the aims, instructions, and a definition of visual imagery as *"the spontaneous formation of visual images or pictures in your mind's eye. Your imagery experience is completely subjective, and it is completely acceptable not to have experienced any imagery at all"*. Survey 1 took approximately 12 minutes to complete. The presentation order of musical stimuli was randomised across participants.

Participants were advised to use good-quality headphones and to have minimal outside disturbance throughout the study. For the three main trials, participants were first presented with the musical excerpt. They were instructed to listen to the whole excerpt and told that on the next page they would be presented with questions regarding their experience of the music. They were advised to pay attention to any visual imagery that they may be experiencing, and the musical characteristics that may have contributed to the imagery (question addressed elsewhere), as they listened. On a first response page, participants were asked to rate the prevalence (the amount experienced) and vividness (the clarity with which it was experienced) of their visual imagery. These ratings were followed by the open-text question asking them to describe the content of their visual imagery (if any). Finally, participants were asked to rate how much they liked the music, and how intense any felt emotional response to the music was.

After this, participants completed the VVIQ, the musical training dimension of the Gold-MSI, then indicated whether they have participated in any activities associated with the visual arts.

Three weeks later, participants completed an almost identical survey, excluding the VVIQ, Gold-MSI, and question about their experience with the visual arts. Certain demographic questions were collected once again, such as an anonymous ID, but also age, and nationality, to improve the chances of confidently being able to match participant data across both surveys. This second survey took approx. 8 minutes to complete.

2.3. ANALYSIS

2.3.1. Thematic analysis of open-text reports of music-evoked visual imagery content

Using Braun and Clarke's (2006) approach, I implemented a thematic analysis on the responses to the open-text question ('*Please describe the content of your visual imagery (if any)*') to identify the prominent themes that emerged in terms of visual imagery in response to the three musical excerpts. This thematic analysis was carried out using Microsoft Excel (Version 16.43) while further quantitative analyses were calculated using R (Version 4.2.3; R Core Team, 2018).

The thematic analysis was carried out by two independent coders to reduce the risk of personal bias. The dataset was analysed as one large unit independent of excerpt emotion with the aim of producing a rich account of each developing theme. In general, to overcome any potential variability in language as a result of participants' wide range of personal and musical backgrounds, I discerned meaning from the data at an explicit level (i.e., a literal interpretation of the text).

The dataset from survey 1 comprised a total of 1,059 reports across the three musical excerpts, while survey 2 comprised a total of 762 reports. Two coders independently examined the dataset and identified Level 3 codes that could encompass key aspects of each description, as well as making note of as many ideas and patterns that may benefit the final model as relevant. Commonalities between specific terms were identified from Level 3 (L3) codes, which were sorted and categorised into distinct groups comprising Level 2 (L2) codes. The suitability of the subthemes in this level was reviewed by the two coders, including whether to collapse redundant subthemes or to divide ones that were too diverse. The L2 codes were finally combined to form higher-order Level 1 (L1) themes. The final structure was discussed by the two coders, confirming that it formed an effective hierarchy that offered a parsimonious overview of the content of the free-form descriptions.

2.3.2. Measuring consistency within and across reports

I used the Jaccard coefficient index, a conservative measure of agreement that enables assessment of the level of overlap between any two lists (0 = no similarity at all, and 1 = perfect similarity):

$$J(A,B) = \frac{A \cap B}{A \cup B}$$

In order to assess consistency, I needed a way of referring to each visual imagery code in a systematic way. Thus, I created a labelling system where all theme levels from the thematic analysis were assigned a numeric label as a unique identifier. To test the efficacy of these labels as a way to sort each participant report, the two coders first used this system to independently label a subset of 60 participant reports (5-6% of total) from the dataset, using L3 codes (the most detailed level of the visual imagery codes). Each individual report was assigned labels pertaining to the content of their descriptions. This first attempt by the coders yielded a Jaccard similarity average of 69% between Coders 1 and 2. Coders discussed any comments or issues that resulted from this effort, including minor reorganisation of the code levels. As a final check, a new subset of 60 reports was analysed, resulting in a version that was a more feasible and intuitive organisation of the themes. The codebook was reviewed and discussed once more by the research team to address any remaining structural issues. Finally, Coder 1 used the labelling system to assign labels to the rest of the datasets for survey 1 and survey 2. The original and coded framework can be found through the Open Science Framework using the following link: https://osf.io/nf4x7/?view_only=081602e5aca94788bb959b48ed8b47ef.

I assessed two types of consistency for each musical excerpt separately: within- and across-participant consistency. Within-participant consistency was computed by comparing the lists of themes emerging from each participant's reports across surveys 1 and 2, and across-participant consistency was computed by comparing the list of themes from each individual participant with the lists from every other individual in the sample, resulting in 352 values per participant for each musical excerpt type (leading to three groups of values) that were then averaged to create a single consistency value per individual per excerpt. With regard to across-participant consistency specifically, individuals were always compared against other individuals' responses that were within the same excerpt type.

In order to retain a precise estimation of the visual imagery types present in the data and its consistency within the sample, the descriptions of those who reported not experiencing any visual imagery or only vague imagery in one or both timepoints were excluded before running the Jaccard analyses. However, those who were coded as having no or vague imagery *were* included if they also reported additional coded experiences that could be used to compute their consistency (i.e., reports describing no or blurred visual imagery, but references to other relevant experiences such as emotional reactions or other imagery types). I also ran the same analyses on the codes only depicting visual experiences (i.e., those under the *Storytelling* higher-order theme, see section *2.4.1*). This was to fully address one of my main aims and research questions regarding the consistency of content of visual imagery more specifically.

2.3.3. Comparing consistency levels and determining the behavioural measures and individual differences that influence them

I assessed the differences in Jaccard coefficient scores between within- and across-participant consistency within each code level (L2 (more broad) versus L3 (more detailed) codes) using linear mixed effects models. The purpose of this was to examine whether consistency levels differ when comparing within and across individuals. To test this, two models were defined, using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages in R, one including L3 consistency values as the dependent variable and the other including L2 values as the dependent variable. In both models, Consistency Type was entered as a categorical fixed effect representing whether the consistency values originate from the Within-Participant or Across-Participant analysis, with participant and musical excerpt entered as random effects in both models.

The same method was used in a subsequent stage of the analysis to test the strength of any relationships between visual imagery consistency values (within- and across-participants) and data on the behavioural experiences of music as well as individual differences. Four models were run, each including within- and across-participant consistency values for each code level (L2 and L3) as dependent variables. In each model, prevalence and vividness of visual imagery, music liking, emotional intensity, VVIQ, musical training, and visual arts participation (*yes* or *no*) were entered as fixed effects, with participant and musical excerpt included as random effects.

2.3.4. Tests of associations and differences between behavioural measures and individual differences

Finally, linear mixed effects models and t-tests were run to ascertain the strength of associations and differences with regard to behavioural ratings (prevalence and vividness of visual imagery, music liking, and emotional intensity) and data on individual differences (VVIQ, musical training, and visual arts participation). To this end, two models were run, one with prevalence as dependent variable and the other with vividness as dependent variable, with music liking, emotional intensity, VVIQ, and musical training included as fixed effects, and participant and musical excerpt as random effects in both models. Further, independent samples t-tests were run to assess differences between those who do and do not participate in the visual arts in terms of their prevalence and vividness of visual imagery ratings.

2.4. RESULTS

2.4.1. Thematic analysis of music-evoked visual imagery

As can be observed in Table 2-2, the thematic analysis conducted on descriptions of visual imagery content resulted in three higher-order themes pertaining to the most prevalent topics in participant reports. As will be observed below, participant descriptions did not always include experiences that were strictly visual in content and often included descriptions related to affect, musical features, and other forms of mental imagery. This fact has been taken into consideration with regard to analyses and conclusions drawn.

A summary of the frequencies of each L2 subtheme is compiled in Figure 2-1 (see Figure A.1 in Appendix A for frequencies of L2 codes across music excerpts, and Figure A.2 in Appendix A for frequencies of L3 codes). I summarise each higher-order theme and their subthemes (in descending order of prevalence) as follows:

Figure 2-1

Occurrence of L2 visual imagery content

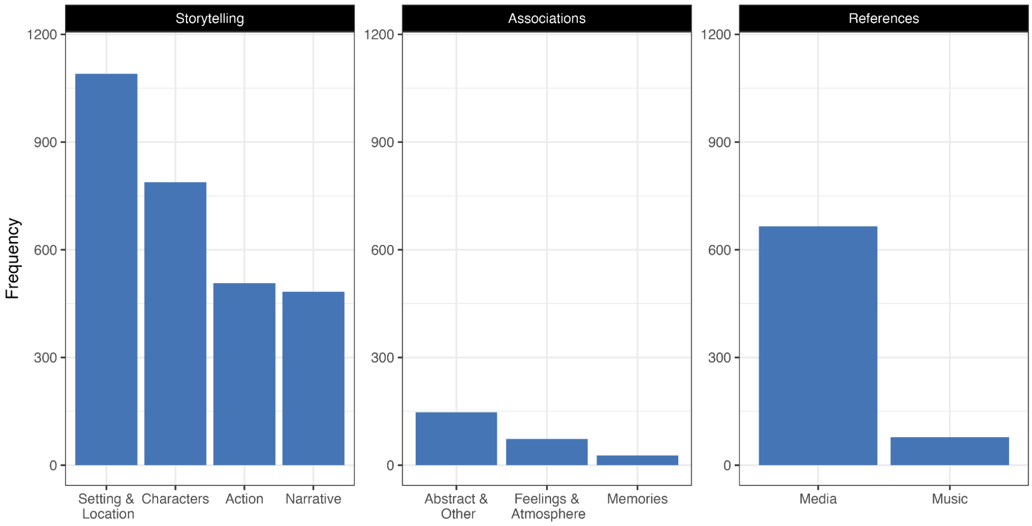


Table 2-2

Thematic framework of prominent higher-order themes (L1) of music-evoked visual imagery content (with L2 and L3 codes in descending order of prevalence). Values included beside L1 codes represent their proportions within the entire framework, whereas the two values beside L2 codes reflect their proportions within their higher-order theme (left) and within the entire framework (right).

Level 1	**Storytelling (74.3%)**				**Associations (6.4%)**			**References (19.3%)**	
Level 2	**Setting & Location (38.0%, 28.2%)**	**Characters (27.5%, 20.4%)**	**Action (17.7%, 13.1%)**	**Narrative (16.8%, 12.5%)**	**Abstract & Other (59.9%, 3.8%)**	**Feelings & Atmospheres (29.6%, 1.9%)**	**Memories (10.9%, 0.7%)**	**Other Media (89.5%, 17.2%)**	**Music (10.5%, 2.0%)**
Level 3	Building	Imagery of Self	Communicative Gestures	Plot Lines	Physical Reactions	Emotional Reaction/Feelings	Autobiographical	Medium	Genre
	Urban	Individuals	Idle/Passive	Hardship	Other Senses	Atmosphere/Mood	Episodic	Genres	Composer
	Nature	Royal	Mental	Conflict & Resolution	Cross-Modal			Specific Movie/Game/Show	Characteristics
	Interior & Exterior Features	Heroic	Interaction	Celebratory				Features	Instruments
	Existing Locations	Evildoer	Musical	Inquisitive					
	Imagined Setting	Civilian	Leisure						
	Time References	(Social) Groups	Gait						
	Season	Musical							
	Weather	Animals/Insects							
	Other Objects	Characteristics							

Storytelling

Storytelling, the first higher-order theme, is characterised by descriptions of narratives involving situations, actions, people, and locations, thus confirming Hypothesis 1. In most cases, descriptions referred to fictional situations involving non-existent characters, although participants sometimes also visualised themselves or other familiar individuals in imagined circumstances. This higher-order theme most explicitly encompasses details that were visualised. This higher-order theme was referred to in 74.3% of the total sample (2,868 times) throughout reports, and comprised four L2 subthemes:

1) *Setting & Location:* describing the locations, interior/exterior, and temporal (and other) details in which the scenes took place. This comprised 38.0% of the *Storytelling* theme (28.2% of total, 1,090 times, e.g., "*It was a **tree on top of a hill***", "*I was in a **museum** on **a day with beautiful weather***")
2) *Character:* referring to the presence of an individual, groups of people, animal, or imagery of oneself within content descriptions. This comprised 27.5% of the *Storytelling* theme, 20.4% of total, 788 times). This subtheme comprised (fictional or real) individuals ranging in background or societal status and includes a lower-level theme describing a range of bodily characteristics (e.g., "*I feel like I'm at the coronation of a future **king***", "***A man** in a **button-down, sleeves** rolled up to mid-arms, sitting in an apartment*")
3) *Action:* referring to actions or activities performed by individuals contained in the descriptions. This comprised 17.7% of the *Storytelling* theme (13.1% of total, 507 times, e.g., "*A couple **walking** on the beach*", "*Man and woman **dancing** slowly*")

4) *Narratives:* referring to plot transitions and prevalent themes portrayed by the content. This comprised 16.8% of the *Storytelling* theme (12.5% of total, 483 times, e.g., themes of celebration or conflict, "*A celebration* ***of a great event*** *with lots of people...*", "*Some scary monsters, and* ***very dynamic action****...*", "***At first,*** *I sort of had an imagery of a battlefield ...* ***then later on*** *it was more sort of fairy-tale like*")

Associations

Associations, the second higher-order theme, occurred 6.4% of the total number of reports (247 times throughout reports). This theme encompasses a mixture of perceptual or sensorial experiences within participants' reports to the music and/or those contained in narrative descriptions of the music, as well as some mention of abstract forms of visual imagery. In all, this higher order theme groups together multimodal and affective experiences contained within the musical experience. It comprised three L2 subthemes:

1) *Abstract & Other:* referring mostly to abstract forms of visual imagery, mental imagery additional to visual imagery, as well as physical experiences in relation to the music (e.g., "*At first, I* ***saw light and bright colour****s and later ... clouds above* ***a red colour***", "*I felt my* ***eyes tremble***"). This subtheme occurred 59.5% of the *Association* theme (3.8% of total, 147 times)
2) *Feelings & Atmosphere*: comprising 29.6% of the *Associations* theme (1.9% of total, 73 times). This subtheme describes any emotional reactions or moods portrayed by the scene or any felt by characters within the narrative or participant themselves (e.g., "***Love and happiness*** *in general*", "*A* ***scary mood, oppressive****, a* ***danger***"). Whilst not

representing visual imagery, its presence within reports was considerably prevalent and constituted a prominent attribute of descriptions of visual imagery experience

3) *Memories:* describing two types of memories that participants may have recalled during their listening experience: autobiographical (e.g., "*It* ***reminds*** *me of the old cartoons*") and episodic (e.g., "*Walking to the stage* ***on my graduation***"). This theme was the least prevalent amongst the other elements in the framework, occurring only 10.9% of the already very small *Associations* theme (0.7% of total, 27 times, e.g., "***My*** *ballet classes*")

References

References, the final higher-order theme, occurred in 19.3% of the total sample (743 times) throughout reports. It is comprised of two L2 subthemes and contains codes referring to details pertaining to the origin of the participant's visual imagery, and generally also includes a mixture of semantic associations (e.g., the music being characteristic of a particular genre) and/or visualised details (e.g., specific instruments or composers). It comprised two L2 subthemes:

1) *Media*: the first and majority subtheme (89.5% of the theme, 17.2% of total, 665 times), comprised references to different types of media (e.g., television, film), references to media genres (e.g., horror, action, film noir), and mentions of any pre-existing media (e.g., "***Animated Disney-style movie***", "*A* ***40s horror movie***")
2) *Music*: comprised references to musical genres (10.5% of the theme, 2.0% of total, 78 times), composers, or instruments that were visualised or described as contributing to the visual imagery (e.g., "*As if I were sitting with tea in my hand next to a famous*

composer like ***Fryderyk Chopin***", "***When the piano started playing*** *I imagined a dark haired male pianist on a black* ***piano****...*")

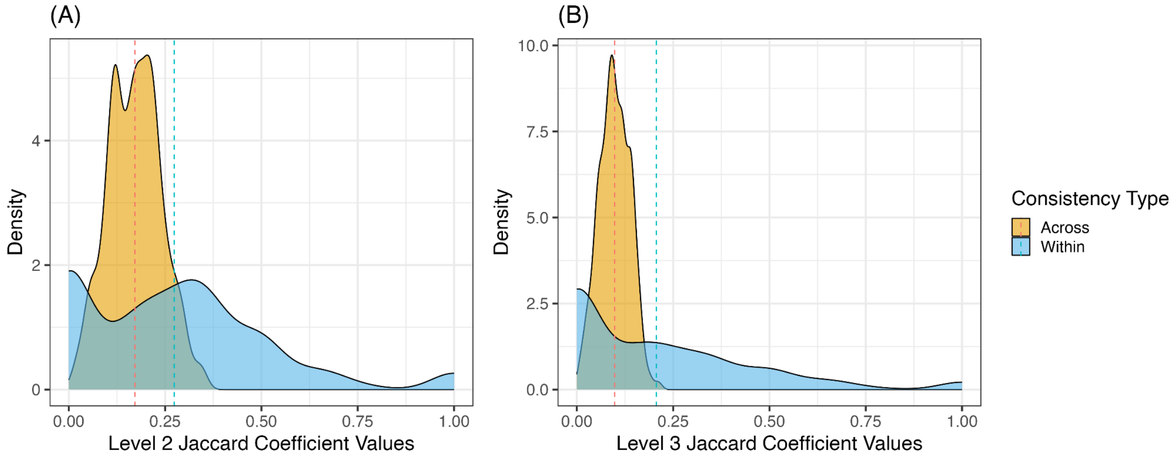
2.4.2. Do respondents show higher consistency with themselves than others?

Table 2-3 provides examples of reports at different (full framework) within-participant consistency boundaries. While reports at 0% reflect no discernible overlap, the 20-50% consistency range reflects minor content overlap in reports between timepoints, the 50-99% consistency range denotes significant content overlap between timepoints, and 100% reflects complete overlap and almost identical responses between timepoints.

Figure 2-2 illustrates the distributions of within- and across-participant consistency (and Figure A.3 in Appendix A demonstrates these distributions as a function of music excerpt type). The graphs display marked differences in range, with the within-participant distribution spanning a wider set of consistency values and additionally revealing a higher proportion of zero levels of consistency, while across-participant consistency appears to span a much narrower value range and peaks at a relatively low level (see also Tables A.1 and A.2 in Appendix A for precise consistency values for within- and across-participant profiles, respectively, at different Jaccard percentage ranges).

Figure 2-2

Comparing density distributions of consistency values of within- (across listening situations) and across-participant (across listeners) groups with mean intercepts. (A) Level 2 codes. (B) Level 3 codes.



In order to test Hypothesis 2, I asked whether there were differences in consistency when comparing an individual with themselves at a later stage, on the one hand, to when comparing a given individual with the remaining sample. The model comparing Jaccard coefficient values for L2 codes between the within- (*Mean* = 27.4%, *Median* = 25%) or across-participant (*Mean* = 17.1%, *Median* = 17.2%) distributions, indeed, confirmed there to be distinguishable differences between the two groups (ß = 0.10, SE = 0.01, $t = 13.00$, $p < 0.001$), with the within-participant group displaying overall greater consistency values than the across-participant group. Additionally, the model predicting coefficient values at the higher granularity L3 codes similarly showed there to be overall differences between the within- (*Mean* = 20.6%, *Median* = 14.3%) and across-participant (*Mean* = 9.8%, *Median* = 9.6%) groups (ß = 0.11, SE = 0.01, $t = 14.24$, $p < 0.001$).

Figure A.4 in Appendix A presents within- and across-participant distributions using only *Storytelling* codes (see also Tables A.3 and A.4 in Appendix A for precise consistency values for these codes), since these, unlike the *Associations* and *References* codes, constitute codes that characterise purely visual imagery experiences. I once more show the two distributions to possess distinct shapes and peaks as previously described for all codes. Again, the model comparing Jaccard coefficient values for L2 codes between the within- (*Mean* = 33.6%, *Median* = 28.6%) or across-participant (*Mean* = 21.5%, *Median* = 21.2%) distributions for *Storytelling* codes confirmed differences between the two groups (ß = 0.12, SE = 0.01, t = 12.06, $p < 0.001$). Finally, the model predicting consistency values at the higher granularity L3 codes also demonstrated overall differences between the within- (*Mean* = 26.1%, *Median* = 16.6%) and across-participants (*Mean* = 11.9%, *Median* = 11.3%) groups (ß = 0.14, SE = 0.01, t = 14.35, $p < 0.001$).

Table 2-3

Example visual imagery excerpts at different bands of within-participant consistency L3 codes.

Survey 1 Excerpts	Survey 1 Codes	Survey 2 Excerpts	Survey 2 Codes	Consistency Level	Reports at this level (% of N)
Akin to the end of a children's fantasy movie	1.1.1, 2.3.1, 2.3.2, 1.4.7	*Gathering of people for a celebration of royalty*	1.1.4, 1.4.7, 1.4.3	0-19.99%	399 (55.4%)
I'm watching a theatre play, and it's starting.	1.4.1, 1.2.5, 1.1.5	*I have imagined that I'm in a big forest to explore it*	1.3.1, 1.1.3, 1.2.6		
I imagined a man playing a piano in an empty room	1.4.2, 1.2.5, 4.1.4, 1.3.1, 1.3.4	*I felt like I was watching someone playing on a piano in a massive room. Beautiful, relaxing...and sort of nostalgic.*	1.2.5, 1.4.2, 3.2.4, 1.3.1, 2.3.1	20-49.99%	208 (28.9%)
The celebration of a victory at the end of a war	1.1.3, 1.1.4	*A hero triumphant return to its home village*	1.4.4, 1.1.3, 1.3.1		
WWI trenches, then galloping horses, then marching soldiers	1.1.1, 1.1.3, 1.2.6, 1.4.9, 1.2.7, 1.4.3	*WWI, American civil war battles, armies marching*	1.1.3, 1.4.3, 1.2.7	50-99.99%	90 (12.5%)
I feel like I'm at the coronation of a future king	1.4.1, 1.1.4, 1.4.3	*I have the impression that I was at the coronation of the king / queen, or at some state event*	1.4.1, 1.4.3, 1.3.1, 1.1.4		
Lonely day at bar/cafeteria. Raining outside. Lonely afternoon.	1.3.1, 2.3.1, 1.3.9, 1.3.7	*Lonely evening at the bar. Raining outside.*	2.3.1, 1.3.7, 1.3.1, 1.3.9	100%	23 (3.2%)
I imagine someone of royalty getting married like for example Prince William and Kate.	1.4.3, 1.1.4	*I imagine a wedding of someone of the royalty.*	1.1.4, 1.4.3		

Note. N = 720

2.4.3. Associations between visual imagery consistency and musical experience

In the next set of analyses, I asked to what extent across-participant consistency (a participant compared with everyone else in the sample) or within-participant consistency (a participant compared with themselves) were associated with participants' experience of the music (visual imagery prevalence and vividness, liking, and emotional intensity) on the one hand, but also individual differences in general visual imagery ability, musical training, and participation in the visual arts on the other (see Table 2-4 for a full summary of the model results).

The model predicting L2 across-participant consistency revealed prevalence (ß = 0.01, SE = 0.00, $t = 2.96$, $p = 0.003$) and vividness (ß = 0.00, SE = 0.00, $t = 2.49$, $p = 0.013$) of visual imagery to be significant predictors of this level of consistency, whereas the model predicting L3 across-participant consistency showed only prevalence to be a significant predictor (ß = 0.00, SE = 0.00, $t = 2.30$, $p = 0.022$). No other behavioural measures or individual differences significantly predicted across-participant consistency in either level.

Conversely, the model predicting L2 within-participant consistency revealed that none of the behavioural measures possessed any predictive influence over this consistency level. However, the model predicting L3 within-participant consistency showed a significant influence of the vividness of visual imagery (ß = 0.02, SE = 0.01, $t = 2.18$, $p = 0.030$). No other behavioural measures or individual differences significantly predicted within-participant consistency in either level.

Table 2-4

Fixed effect estimates of behavioural and individual difference measures predicting within- and across-participant consistency.

	Across Consistency L2				**Across Consistency L3**				**Within Consistency L2**				**Within Consistency L3**			
	ß	**SE**	***t***	***p***	**ß**	**SE**	***t***	***p***	**ß**	**SE**	***t***	***p***	**ß**	**SE**	***t***	***p***
Intercept	0.11	0.02	6.32	**< 0.001*****	0.09	0.01	7.84	**< 0.001*****	0.09	0.06	1.40	0.163	0.10	0.06	1.61	0.110
Imagery Prevalence	0.01	0.00	2.96	**0.003****	0.00	0.00	2.30	**0.022***	0.01	0.01	1.18	0.238	0.01	0.01	1.20	0.230
Imagery Vividness	0.00	0.00	2.49	**0.013***	0.00	0.00	1.24	0.214	0.01	0.01	1.42	0.155	0.02	0.01	2.17	**0.030***
Music Liking	0.00	0.00	0.62	0.534	–0.00	0.00	–0.16	0.869	0.01	0.01	0.94	0.356	–0.01	0.01	–1.30	0.193
Emotional Intensity	0.00	0.00	1.75	0.080	–0.00	0.00	–0.62	0.534	0.01	0.01	1.10	0.274	0.01	0.01	0.77	0.440
VVIQ	0.00	0.00	0.44	0.658	–0.00	0.00	–0.68	0.495	0.01	0.01	0.47	0.637	–0.01	0.01	–0.41	0.684
Musical Training	–0.00	0.00	–1.41	0.161	–0.00	0.00	–0.56	0.576	–0.00	0.01	–0.30	0.768	0.00	0.01	0.62	0.534
Visual Arts	0.00	0.01	0.72	0.470	0.00	0.00	0.36	0.716	–0.03	0.02	–1.27	0.204	–0.01	0.02	–0.50	0.618

Note. *** < 0.001, ** < 0.01, * < 0.05. Abbreviations: L2 = Level 2, L3 = Level 3.

2.4.4. Prevalence and vividness of music-evoked visual imagery and links between visual imagery, emotional intensity, liking, and individual differences

Table 2-5 presents a summary of visual imagery prevalence and vividness ratings divided and aggregated by musical excerpt. These values demonstrate high proportions of visual imagery prevalence and vividness across all musical excerpts, that persist even when one considers the individual excerpt types. Higher prevalence and vividness levels of visual imagery can be seen in response to the Fearful excerpt than the Happy and Tender excerpts, which both exhibit almost equal proportions.

Table 2-5

Proportions (and frequencies) of the prevalence and vividness of visual imagery for and across the musical excerpts.

	Happy		**Tender**		**Fearful**		**Overall**	
	No VI	**At least mild VI**	**No VI**	**At least mild VI**	**No VI**	**At least mild VI**	**No VI**	**At least mild VI**
Prevalence	8.8% (31)	90.7% (320)	9.6% (34)	90.4% (319)	5.1% (18)	94.6% (334)	2.5% (9)	97.5% (344)
Vividness	10.9% (38)	89% (314)	13.6% (48)	86.1% (304)	5.9% (21)	93.2% (329)	5.4% (19)	94.6% (334)

Note. Abbreviations: VI = visual imagery. “No VI” pertains to individuals who provided the lowest rating, whereas “At least mild VI” refers to ratings of 2 and upwards. Values in brackets indicate the sum of reports within that category. Combined sum of reports for or across excerpt types for each rating that do not equate to the total number of participants (N = 353) are due to missing ratings data.

Confirming Hypotheses 3, 4, 5, and 6, the model predicting prevalence of visual imagery demonstrates that emotional intensity (ß = 0.70, SE = 0.04, $t = 15.83$, $p < 0.001$), music liking (ß = 0.25, SE = 0.04, $t = 5.53$, $p < 0.001$), and the VVIQ (ß = 0.48, SE = 0.08, $t = 5.95$, $p < 0.001$) were all highly significant predictors. Contrastingly, musical training (ß = 0.02, SE = 0.04, $t = 0.65$, $p = 0.517$) did not significantly predict visual imagery prevalence.

Similarly confirming my hypotheses, the model predicting vividness of visual imagery similarly showed emotional intensity (ß = 0.70, SE = 0.04, $t = 15.25$, $p < 0.001$), music liking (ß = 0.21, SE = 0.04, $t = 4.64$, $p < 0.001$), and the VVIQ (ß = 0.54, SE = 0.08, $t = 6.45$, $p <$

0.001) to be highly significant predictors, whereas, once again, musical training was not (ß = –0.02, SE = 0.04, $t = –0.45$, $p = 0.655$).

In response to the question regarding experience with activities in the visual arts (VA), 20.1% (n = 71) reported that they participate in activities associated with the visual arts, which included activities such as painting, photography, and graphic design. With regard to the averaged excerpt ratings between those who do and do not (NVA) participate in the visual arts, independent samples t-tests showed that those who participate in the visual arts reported significantly more visual imagery prevalence (*Mean*-VA = 4.63, *Mean*-NVA = 3.99, $t(119.4) = 3.78$, $p < 0.001$) and more vividness (*Mean*-VA = 4.30, *Mean*-NVA = 3.74, $t(124.1) = 3.37$, $p < 0.001$) than those who do not, supporting Hypothesis 7. There were also significant differences with regard to each of the individual musical tracks, with individuals who do take part in the visual arts reporting higher prevalence and vividness than those who do not (see Tables A.5 and A.6 in Appendix A for full results).

2.5. DISCUSSION

The aims of Study 1 were multi-fold. The main aim was to derive the most prominent themes present in listeners' descriptions of their music-evoked visual imagery and to investigate the extent to which listeners exhibited consistency in their visual imagery reports within themselves (within-participants) and with the whole cohort (across-participants). Further to this aim, I sought to explore whether the prevalence and vividness of visual imagery, music liking and emotional intensity, as well as individual differences in general visual imagery ability, musical training, and participation in the visual arts may be associated with patterns of consistency levels shown. However, other important aims were to replicate previous findings of a link between visual imagery and emotion (Day & Thompson, 2019; Vuoskoski & Eerola, 2015), aesthetic appeal (Belfi, 2019), and to explore the influence of general visual imagery ability, musical training, and participation in the visual arts on music-evoked visual imagery experience.

In sum, I found (a) that storytelling is the most common form of visual imagery during music listening, (b) individuals are more consistent with themselves (in terms of their visual imagery content) than when compared to other listeners, (c) evidence for all-round modest consistency levels, (d) confirmation of links between visual imagery, emotional intensity, and aesthetic appeal, and (e) evidence for a nuanced role of individual differences on music-evoked visual imagery experience in terms of prevalence, vividness, and consistency. Further to these findings, it was ascertained that visual imagery experience averaged across the three music stimuli was very prevalent in my sample with about 97.5% of participants reporting experiencing at least mild levels of visual imagery, and about 94.6% reporting their visual imagery to be at least a mildly vivid experience (similar proportions were also found across the individual music excerpts).

2.5.1. Storytelling is the most common form of visual imagery during music listening

The thematic analysis of the open-text question revealed that (in support of H1) story-making was a very pervasive aspect of visual imagery experience. Indeed, the first higher-order theme *Storytelling* encompassed descriptions of visual imagery ranging from locations to individual characters and was present in over 74% of reports. This finding is in line with the idea that individuals are prone to imagining narratives in response to music (Margulis, 2017; Margulis, Wong, et al., 2022), and the findings support the notion that this may significantly occur in the visual domain. In addition to showing that visualising story-making is a prevalent aspect of music listening, I was also able to offer insights into the relative prevalence of certain types of visual imagery such as the prominence of details pertaining to setting and location, followed by characters, and the actions they carried out.

I coded two further themes that accompanied my sample's visual imagery reports, *Associations*, comprising abstract visual imagery, emotion, and memories, and *References*, comprising codes regarding media comparisons, instruments, and composers. Codes underlying these themes were present in over 10% of reports and support past arguments that music listening is a multimodal experience (Deroy, 2020; Nanay, 2018) that involves not just visual experiences but also physical, affective, and semantic connotations. In other words, in addition to reports suggesting that certain visual imagery was in fact 'seen', descriptions were also often accompanied by remarks of cross-modal aspects spanning the remaining senses, aesthetic evaluations, and music's emotional power. These findings highlight the considerable prevalence of semantic associations found within listeners' descriptions in response to the music, even when explicitly instructed to focus their attention on visual imagery. Such patterns indicate that narrowing one's research aims on just the visual components of a listeners' experience could lead to overestimations of its occurrence as well as overshadows its potential unique links to other semantic associations formed in response to the music; one recent study by Cespedes-Guevara and Dibben (2022) showed that a considerable portion of listeners' reports of what went through their minds while listening comprised an array of semantic, personal, as well as visual experiences.

2.5.2. Participants show all-round low consistency levels but greater levels for within- than across-participants

I investigated visual imagery consistency using the Jaccard coefficient index, a measure of overlap between two lists. First, I assessed within-participants consistency by comparing each individuals' content from surveys that were administered three weeks apart. Here, I found that, overall, consistency levels tended to cluster around 20%, indicating that listeners tended to refer to almost a fifth of the same visual imagery features in both instances. The analysis of across-participant consistency yielded similarly low results when comparing individual participants with the rest of the sample: approx. 7-13% for L3 codes and 13-22% for L2 codes.

At first glance, the findings show notable differences with a recent investigation by Margulis et al. (2022). The authors reported high levels of consistency in their sample with regard to musical narrative engagement that they suggest was, for the most part, determined by cultural experience. However, it is important to point out both the differences in what was being reported on (visual imagery vs. imagined narrative) and how consistency is estimated in the two papers. Indeed, I opted to use the Jaccard coefficient index on lists of themes, a method which takes the presence of individual cases into consideration. In contrast, Margulis and colleagues' approach focuses on the respondents' text as a whole, using a cosine similarity as a weighted index to assess semantic ‘closeness’ between portions of text based on their orientations on a multi-dimensional space. Approaches used by other studies to address similar questions have also included frequency-based approaches (i.e., summing occurrences of certain content, Dahl et al., 2022). Critically, the current research aimed to assess consistency with a focus on the presence or absence of particular themes and topics. By allowing analysis of consistency on the basis of themes and topics, my approach provides a way to estimate consistency levels with a focus on particular aspects of content. However, it would be beneficial to consider the advantages of applying a weighted approach (as Margulis and colleagues have done) when scrutinising visual imagery content in combination with the current study’s methods of assessing discrete overlap. Incorporating both topic overlap as well as weighted similarity might provide insight on the significance of specific types of cross-modal content evidently present (as is seen within the thematic framework) in reports of music-evoked visual imagery.

In any case, I was able to confirm H2 about how within- and across-participant consistency would differ from each other, supporting Cross’ (2011) aforementioned idea on the subjectivity of music. Specifically, I show that participants possess greater consistency within themselves across two time-points than consistency with other listeners. Critically, this type of behaviour is not so different to what researchers have described as cross-modal correspondences. Deroy and Spence (2016) explain that the sensory connections or cross-modal matching are not only pervasive in everyday life but also remain quite consistent over time, which could be down to environmental regularities or contextual associations.

In a final check, I aimed to re-assess consistency by only including *Storytelling* codes. This was viewed as an important next step to (a) confirm that it was indeed pure visual imagery (i.e., codes where participants report seeing images) that was mostly leading to the observed consistency levels above, and (b) aid comparison with previous work on narrative consistency during music listening (Margulis, Wong, et al., 2022). I was able to show that the main conclusions that I drew (regarding comparison of within- and across-participant consistency profiles across code levels) all continued to hold even when looking at this reduced set of codes.

2.5.3. Relating visual imagery consistency to behavioural ratings and individual differences

Next, I explored whether there was a predictive influence of the behavioural measures (prevalence, vividness, music liking, and emotional intensity) and individual differences (general imagery ability, musical training, and participation in the visual arts) on within- and across-participant consistency. There were links between across-participant consistency and the prevalence and vividness of visual imagery (albeit only L2 for vividness). Interestingly, within-participant consistency was only predicted by vividness and only for L3 codes.

The generally positive relationships found between (across- and within-participant) consistency and prevalence and vividness of visual imagery clearly indicate that the qualitative nature of listeners' visual imagery influences how consistent they are within themselves and with others. While it may not be appropriate to speculate on the nuanced differences seen with regard to levels of consistency, my results suggest that music that is able to induce highly vivid visual imagery tends to produce largely similar content across listeners. Given that I had explicitly chosen to present musical stimuli capable of inducing distinct types of emotion that likely vary in their compositional techniques, I propose that the unique acoustic features of the different excerpts could have influenced how vividly listeners experienced their visual imagery; an idea in line with literature that has highlighted features such as contrast in promoting narrative thinking (Margulis, 2017), although is an area of research that is generally in need of much more insight in order to properly ascertain its influence over visual imagery formation (Juslin, 2019).

Further, the analyses show that participation in the visual arts does not predict within- or across-participant consistency at either level of granularity, aligning with the idea that those who participate in the visual arts could experience visual imagery at a high enough frequency, vividness, and creative freedom to be inconsistent in content.

2.5.4. Visual imagery, emotion, and aesthetic appeal

In line with previous work which has linked emotion induction with visual imagery formation (Balteș & Miu, 2014; Hashim et al., 2020; Juslin, 2013; Juslin & Västfjäll, 2008; Taruffi & Küssner, 2019; Vroegh, 2018), there were significant associations between the prevalence and vividness of visual imagery and emotional intensity, whereby the amount of visual imagery and its vividness were positively related to the intensity of emotions felt, confirming H3.

However, the results were not able to speak to the nature of the relationship between emotion and visual imagery. Indeed, there has been increasing debate regarding the directionality of this relationship – recent studies suggest that it is the emotion that is first felt that then leads to visual imagery experience (Day & Thompson, 2019), whereas others propose a more complex and mutually beneficial interlink between music-evoked visual imagery and emotional induction (Vroegh, 2018). In at least one study, it has been shown that manipulating listeners' experience of visual imagery by attempting to hinder it led to a mild suppression of reported induced emotion (Hashim et al., 2020).

Even though the results cannot speak directly to the directionality of the relationship between visual imagery and emotion induction, they extend findings in a number of useful ways. Although the analyses showed that emotional intensity did not predict L3 across-participant consistency, the pattern for L2 consistency was just shy of significance, hinting at the idea that a minor part of what determines high consistency across individuals may be a shared high level of emotion. Future studies could further probe this link by assessing the extent to which emotional valence is a factor driving similarities in visual imagery reports. Secondly, with the thematic analysis, a lower-level code dedicated to outlining emotional experiences reported in visual imagery descriptions was obtained. Indeed, emotional experiences, whether felt or perceived within the imagery, became a prominent aspect of my categorisations and was highly intertwined in descriptions of different visual imagery scenarios (Küssner & Eerola, 2019); this being further shown in the fact that it was the second most commonly occurring non-visual subtheme of the *Associations* higher-order theme. Again, while this does not offer evidence into the causal relationship between visual imagery and emotion induction, it speaks to the strong link visual imagery has with emotional experiences.

Finally, in line with previous work that showed visual imagery vividness to be strongly linked with music's aesthetic appeal (Belfi, 2019), and in support of H4, I was also able to reflect this finding between prevalence and vividness of visual imagery and music liking responses, whereby liking was positively associated with visual imagery prevalence and vividness. However, this relationship was weaker than seen between visual imagery and emotion, suggesting that it may not be a leading determinant of music-evoked visual imagery experience. Here as well, I was unable to provide full insight into the causal relationship between visual imagery and liking. However, the generally weak associations found draw into doubt the idea that aesthetic appeal may play a major role in influencing visual imagery experience. With regard to what may cause a high amount of music-evoked visual imagery, one might expect that aspects of acoustic and musical features like melody and harmony

(Juslin, 2019), structural features like tempo (Herff et al., 2021), or even interindividual factors like trait empathy (Taruffi et al., 2021), may play as important if not more of a role than liking.

2.5.5. The roles of individual differences on music-evoked visual imagery

I further explored whether music-evoked visual imagery ratings were predicted by generally occurring visual imagery levels. In contrast with previous assessments (Hashim et al., 2020), it was found that general visual imagery showed a moderate positive predictive association with music-evoked visual imagery, supporting H5 and suggesting that general and music-evoked visual imagery may be somewhat interrelated. This finding is partly in line with findings by Küssner and Eerola (2019) who showed a positive, albeit small, correlation between imagery vividness and general visual imagery. Nevertheless, due to the inconsistencies found across studies, these results warrant further investigations into the processes that may underlie the experience of general as well as music-evoked visual imagery.

Furthermore, musical training showed no associations with prevalence and vividness of visual imagery ratings (extending H6 to show that the link is not only weak but in fact non-existent) and is in contrast with past observations that music-evoked visual imagery was affected by training due to potential functional benefits (Küssner & Eerola, 2019). This finding however may not come as such a surprise, as several examples of past literature have similarly found no differences between musicians and non-musicians in their visual imagery abilities (Talamini et al., 2022), instead finding superior involvement of other mental imagery experiences in response to music listening (for instance, auditory, Bishop et al., 2013; Keller & Appel, 2010; and kinaesthetic, Clark & Williamon, 2012; Di Nuovo & Angelica, 2015).

Further, in support of H7, I found that those who participate in activities associated with the visual arts provided significantly higher prevalence and vividness of visual imagery ratings, both in terms of aggregated and individual music track responses. This is unsurprising, as it is evident that various artists (visual, musical, etc.) use imagery in their creative processes to stimulate the creation and performance of their art (Rosenberg & Trusheim, 1989). In the context of music, mental imagery is even found to enhance the perceived creativity of music compositions (Wong & Lim, 2017). Thus, although the general chain of causality is unclear, increased participation within various art modalities (in this case, visual) may increase visual imagery engagement with music.

2.5.6. Implications, limitations, and future directions

With Study 1, I have presented a novel methodological approach to probing the content of music-evoked visual imagery, a method that I hope will be adopted by future studies seeking to develop the knowledge on the topicality of visual imagery content. The ephemeral nature of visual imagery makes it difficult to measure and draw decisive conclusions. Using the approach of quantifying prevalent themes, I was able to corroborate past notions of narrative thinking in the context of music listening and confirm that a large portion of it can be visual in nature.

One important limitation to consider is that the visual imagery content descriptions provided in survey 2 could have been subject to demand characteristics. On the one hand, listeners may have only reported new visual imagery that was not experienced during the first listening instance; on the other, it is possible that some may have felt inclined to report similar visual imagery to that they recalled experiencing in survey 1. Further research on the topic should aim to take this issue into consideration.

Some literature has emphasised the positive effects that eye closure may have on visual imagery experienced in response to (e.g., Hashim et al., 2020; van den Hout et al., 2011; Vredeveldt et al., 2011), so much so that current imagery-based therapies adopt eye closure as a way to enhance the benefits of rehabilitation (Fachner et al., 2019). One study that specifically compares the facilitatory effects of eye opening or closure on participants' reported visual imagery experience found that eye closure led to markedly higher visual imagery vividness as well as content (Herff et al., 2022). The current design did not offer specific instructions on whether participants should listen to each musical excerpt with eyes open or closed, thus this action was free to vary across the sample. Given the dramatic influence that the change in instruction can have on visual imagery experience, especially in the context of music listening, such an instruction would be important to include in future research hoping to enhance the experience of visual imagery.

The design of the current study required participants to provide unrestricted reports on the content of their visual imagery to music, implicating their ability to verbalise their visual imagery experiences (i.e., constructing their narrative to music, Margulis, 2017). It is generally well evidenced that music and language are reflected by overlapping electrophysiological correlates (Carrus et al., 2013; Koelsch, 2009; Koelsch et al., 2005; Kraus & Slater, 2015; Maidhof & Koelsch, 2011; Tillmann, 2012), but some studies identify differences in this regard (e.g., between males and females, Koelsch et al., 2003, 2003; Kramer et al., 1988; Steinbeis & Koelsch, 2008a), and as a function of musicianship (Aleman et al., 2000; Altenmüller et al., 2002; Bangert & Altenmüller, 2003; Brochard et al., 2004; Klein et al., 2015; Lotze, 2013; Trainor et al., 2009). Neuroscientific studies on music-evoked visual imagery are only beginning to emerge (see Fachner et al., 2019; Chapters 3 – 5 of this thesis, for first evidence of neural signatures). However, I suggest that future studies may seek to combine approaches like those taken in the current study with emerging insights into neural underpinnings in order to advance knowledge of both the brain and visual imagery during music listening.

What musical characteristics are more likely to result in music-evoked visual imagery? While this concept has been minimally addressed, Margulis (2017) found that one potential contributing factor in leading listeners to narratively engage with music is musical contrast (sudden and unexpected changes in musical events). Juslin (2019) proposes that musical features, such as predictability and repetition, may be driving forces in leading listeners to form visual imagery to music. In any case, the musical features that may link to specific types of visual imagery is, to date, a vast and unanswered question.

Finally, future investigations may consider extending the time between administering the first and second survey to assess the endurance of visual imagery experience more effectively. A yet more comprehensive approach would involve presenting additional administrations of the survey: this is to assess potential modulations more systematically in terms of how levels of within-person consistency change over time.

2.5.7. Conclusion

In Study 1, I presented a detailed investigation into the visual imagery content that listeners experience in response to music. I show that visual imagery is a highly prevalent aspect of individuals' listening experience, with storytelling being particularly prominent. I also demonstrate the idiosyncrasies of listeners' content consistency by showing that they were, on average, relatively consistent with themselves across timepoints, in contrast to when compared with other listeners. The ease with which music appears to elicit visual imagery offers further support for the connection between music and language processing with regard to listeners' inclination to derive meaning from the music. I anticipate that this study will set a precedent for further studies to develop and hone our understanding of the inherent visual imagery qualities experienced during music listening. In the following chapter, I use EEG to provide neuroscientific evidence for the relationship between activation in visual brain areas and two different types of music-evoked visual imagery, static and dynamic.

CHAPTER 3. THE OSCILLATORY PROFILES OF STATIC AND DYNAMIC MUSIC-EVOKED VISUAL IMAGERY

This chapter presents the results of Study 2, a neuroscientific investigation of music-evoked visual imagery. Music-evoked visual imagery is said to contain a broad variety of content, ranging from abstract shapes to dynamic scenes, and has been linked to visual and motor processing areas of the brain. Forty-two participants listened with closed eyes to twenty-four excerpts of music, while a 15-channel electroencephalography (EEG) was recorded, and, after each excerpt, rated the extent to which they experienced static and dynamic visual imagery. The results show both static and dynamic imagery to be associated with posterior alpha suppression (especially in lower alpha) early in the onset of music listening, while static imagery was associated with an additional alpha enhancement later in the listening experience. With regard to the beta band, the results demonstrate beta enhancement to static imagery, but first beta suppression before enhancement in response to dynamic imagery. I also observed a positive association, early in the listening experience, between gamma power and dynamic imagery ratings that was not present for static imagery ratings. Finally, the study offers evidence that musical training may selectively drive effects found with respect to static and dynamic imagery and alpha, beta, and gamma band oscillations. Taken together, the results of Study 2 show the promise of employing neuroscientific methods for examining visual imagery and its contents. My study also highlights the relevance of future work seeking to study the temporal dynamics of music-evoked visual imagery.

3.1. INTRODUCTION

3.1.1. Comparing electrophysiological correlates of static and dynamic visual imagery

Given evidence that motoric aspects of visual imagery may implicate additional brain areas, an interesting question is how the electrophysiological correlates of static and dynamic forms of visual imagery compare. In one EEG study testing the possibility that pure visual motion imagery can be used as a tool for brain computer interfacing (Sousa et al., 2017), participants were asked to imagine a dot in three modes: static, moving in two opposing directions, and moving in four opposing directions. Compared to observing a static dot on a screen, observing a moving dot led to a greater decrease of alpha levels in posterior brain areas (parietal, parieto-occipital, and occipital areas) supporting the role of visual cortices in visual imagery. However, imagery of the moving dots (compared to the static dot) was also characterised by greater alpha in frontal as well as a decrease of beta activity in fronto-central channels. The authors accounted for their findings of increased frontal alpha to visual motion imagery (compared to static imagery) on the basis of frontal alpha's purported role in tasks with high internal processing demands (e.g., working memory and creative thinking), and based on the association of frontal alpha with reduced external processing and increased task complexity (Cooper et al., 2003; Klimesch et al., 2007; Sauseng et al., 2005; Schomer & Silva, 2012).

Other studies have reported and interpreted beta band involvement in visual imagery even though no solid conclusions have been drawn. Villena-González et al. (2018) found an enhancement of beta power during a task where participants attended to a series of beep tones, following an instruction to visually imagine anything they wanted. They suggested that the presence of beta band activity may indicate the cross-modal processing of the visual imagery and auditory task. In contrast, beta power suppression has been associated with imagination of complex movement in a number of studies (e.g., Menicucci et al., 2020; Zabielska-Mendyk et al., 2018). In the study by Menicucci et al. (2020), parallels were reported between the profiles of beta and alpha band amplitude, with greater beta suppression in fronto-central and centro-parietal areas (along with alpha suppression in fronto-central areas) during a visual motor imagery task. Here, it is also relevant to note studies that emphasise a rebound of beta power – suppression followed by enhancement – after both real and imagined movements (Neuper & Pfurtscheller, 1996; Salmelin et al., 1995). Indeed, one possibility is that both beta suppression and enhancement effects may be expected in dynamic imagery depending on whether motor-associated imagery is being or, conversely, has just been experienced.

In sum, research to date suggests that while visual imagery involves visual cortices, particularly with respect to suppression of alpha activity, incorporating complex features such as motion may engage additional motoric processes in other areas of the brain (for a meta-analysis of the neural correlates of motor imagery, see also Hardwick et al., 2018). In other words, research corroborates the idea that static and dynamic visual imagery experiences may be neurally dissociable. Here, I suggest that music, with its capacity for inducing both static and dynamic forms of visual imagery, may be a useful stimulus for throwing light on this hard-to-grasp phenomenon.

3.1.2. Music-evoked static and dynamic visual imagery

As mentioned in earlier chapters, visual imagery is a common experience during music listening (Küssner & Eerola, 2019; Vuoskoski & Eerola, 2015) with the average latency of music-evoked visual imagery reported as being about 12 s after music onset (Day & Thompson, 2019), and with music-evoked imagery evidenced to have a wide breadth of content (Dahl et al., 2022; Küssner & Eerola, 2019; Taruffi & Küssner, 2019; Vuoskoski & Eerola, 2015).

Indeed, studies requiring participants to describe details of their music-evoked visual imagery experiences have shown how visual imagery can incorporate motoric aspects. For instance, in one recent study by Dahl et al. (2022), a content analysis of music-evoked visual imagery descriptions showed that Movement and Events was a prominent category of listeners' experience. Similarly, Küssner and Eerola (2019) reported on how visual imagery can vary from static scenes to fast changing storylines, while other work has cited dynamic forms of imagery in association with perceived motion and metaphor in musical contexts (Eitan & Granot, 2006; Johnson & Larson, 2003). Within this literature on the cross-modal experience of music, Zhou et al. (2015), for instance, showed that music can express a sense of movement as a function of acoustic parameters such as pitch range and intensity. Further, other authors have emphasised that contours in melodic lines and forces evoked by changes in tempo are what may drive mental imagery of movement (Eitan & Granot, 2006).

However, despite the relevance of using music listening as a vehicle to study the neural correlates of the content of visual imagery (for a general overview of neuroscientific measures of music and mental imagery, see Belfi, 2022), only one such study exists: Fachner et al. (2019) recorded EEG during a guided imagery and music (GIM) session – where GIM involves the induction of visual imagery in response to a specialised GIM soundtrack – from both a therapist and client simultaneously. Comparing moments of interest (characterised by imagery) to moments of non-interest, they found greater posterior alpha suppression during moments of visual imagery formation.

The findings from Fachner and colleagues are promising because they are in line with previous work on visual imagery. However, as that study examined only one participant and as it focused primarily on examining power in the alpha band, it is clear that further work is needed. Indeed, while it is likely that visual imagery to music activates largely similar brain areas as visual imagery to non-music-related stimuli, and while there is preliminary evidence that occipital alpha suppression should be a key signature of interest, it seems relevant to ask whether different forms of music-evoked visual imagery – here, static versus dynamic imagery – may be reflected by different neural patterns.

3.1.3. The current study

In Study 2, the aim was to throw light on the extent to which static and dynamic music-evoked visual imagery may be associated with differing neural patterns with respect to three regions of interests (frontal, centro-parietal, and parieto-occipital) and three frequency bands of interest (alpha [8–13 Hz], beta [14–30 Hz], and gamma [30–45 Hz]). Based on evidence that visual imagery tends to occur within the first 12 s of listening (Day & Thompson, 2019) and in light of evidence that contrasting effects may occur within the same oscillatory frequency bands (e.g., as a function of whether imagery is ongoing or recently completed; Neuper & Pfurtscheller, 1996), I also explored how imagery-related neural activity differed in key phases of the listening experience (the first and second halves of the piece).

Participants were instructed to listen, with closed eyes and while EEG was recorded, to twenty-four excerpts of music that had been shown in a previous study to induce joyful, neutral or fearful emotions in the listener (Koelsch et al., 2013). After each excerpt, they rated – on a continuous scale – the amount of static and dynamic visual imagery experienced in response to the music. In line with previous research, in general (Cooper et al., 2003; Drever, 1955; Gale et al., 1972; Salenius et al., 1995; Williamson et al., 1997; Xie et al., 2020) and in the context of music listening (Fachner et al., 2019), I expected to see a negative relationship between the amount of music-evoked visual imagery reported and alpha band activity particularly in the parieto-occipital area of the brain, reflecting enhanced neural firing in visual areas during imagery.

Critically, I also predicted, in line with past findings regarding beta activity during motor and visual-motor processing (Menicucci et al., 2020; Zabielska-Mendyk et al., 2018), that dynamic imagery may involve both beta suppression (desynchronisation) and enhancement (due to rebound effects), where suppression reflects motor processing, and enhancement reflects (a) a rebound of beta power following such processing (Neuper & Pfurtscheller, 1996), and (b) the potential cross-modal processing of visual imagery and the auditory listening task (Villena-González et al., 2018). Finally, I predicted that I would find parieto-occipital gamma enhancement, reflecting increased neural firing, in response to visual imagery generally (in line with previous results examining vivid spontaneous visual imagery in single case studies, Lehmann et al., 2001; Luft et al., 2019), but potentially in response to dynamic imagery more specifically (in line with past relationships found between motor imagery tasks and gamma enhancement, De Lange et al., 2008; Sepúlveda et al., 2014).

3.2. METHODS

3.2.1. Participants

Forty-three participants took part in the experiment. One participant was excluded due to extreme values that could not be accounted for. This resulted in forty-two participants, aged 19-39 years (M = 28.85, SD = 4.85; 27 female, 15 male; note that age data was missing from one participant), being included in the analyses. About 95 % (40 of 42) of participants reported no hearing issues. One participant mentioned having a hearing aid or implant (i.e., corrected hearing that should not pose any issues for the data), and the final participant provided no response but reported that the musical selection was pleasant and varied and that they listened to all musical stimuli ‘pretty attentively’. Thus, these two participants were retained for further analyses.

3.2.2. Ethics statement

Ethical approval (Ref. 2017-32) for the research was granted by the Ethics Committee of the Institute of Psychology at Humboldt-Universität zu Berlin, Germany. All participants provided written consent to be included and were given monetary compensation or course credit for their time.

3.2.3. Materials and stimuli

Twenty-four musical excerpts that had been shown to induce joyful, fearful, and neutral emotions were chosen (eight excerpts per emotion) for the listening task. They were obtained from a set of stimuli used in a previous study by Koelsch et al. (2013).

The joyful excerpts consisted of CD-recorded pieces derived from a variety of musical styles and genres (classical, South American and Balkan music, Irish jigs, jazz, reggae). The fearful excerpts were obtained from soundtracks of suspense movies and video games. Their fearful qualities were further enhanced by creating two copies from each original excerpt: one copy pitch-shifted a semitone upwards and the other shifted a tritone downwards. The original excerpt and both copies were merged into a single wav file. The neutral excerpts comprised sequences of isochronous tones selected at random from a pentatonic scale, which were set using high quality natural instrument libraries from Ableton (https://www.ableton.com/en/) to ensure ecological validity.

The 7-item Musical Training subscale of the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) was included to gauge prior training in music. Examples of items include ‘I spend a lot of my free time doing music-related activites’ and ‘I would not consider myself a musician’. Each item was rated on a 7-point Likert scale.

3.2.4. Procedure

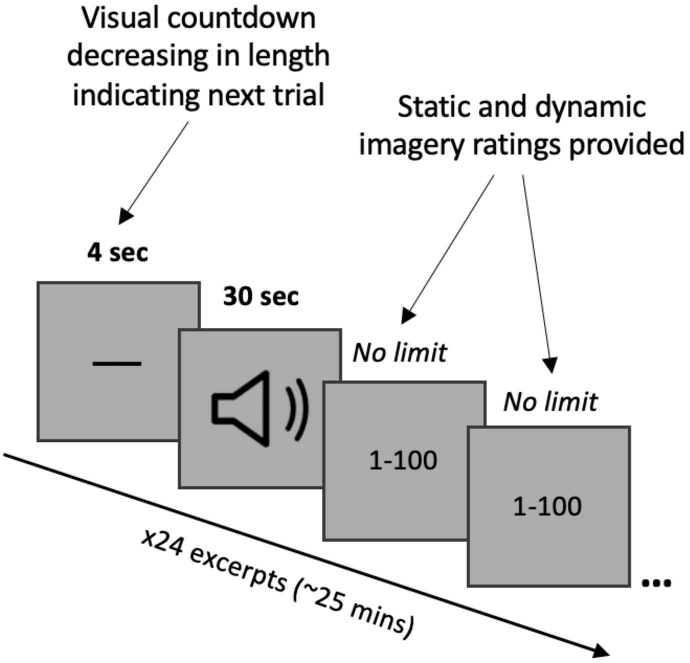
The experiment consisted of two main 28-trial counterbalanced blocks, consisting of 4 practice trials and 24 experimental trials. In one block, participants provided self-report ratings related to visual imagery (analysed here and discussed exclusively henceforth), and in the other, they provided ratings regarding emotion-related experiences, which will be analysed and discussed elsewhere. In both blocks, participants were presented with the same set of twenty-four excerpts (i.e., two repetitions of the musical stimuli). Note that while this means that half of the participants (who were presented with the visual imagery block second in the counterbalancing order) provided their imagery ratings in a second round of listening to the stimuli, this should not affect the conclusions I seek to draw with this study.

Each main block took approximately 25 mins to complete. During the experimental trials, participants listened to the 24 30-s musical excerpts aloud from two speakers set to a volume they found comfortable and, after each excerpt, were instructed to rate their static and dynamic visual imagery experience (‘As soon as the piece of music has finished, please open your eyes and rate how strongly it evoked still and moving images in your mind's eye’) using a visual analogue scale from 0 to 100. Participants always rated their static imagery experience first, followed by their dynamic imagery experience. The static imagery rating ranged from 0 = “Did not trigger any still images in me at all” to 100 = “Triggered a lot of still images in me” (translated from German, see Table B.1 in Appendix B for the original anchor points used). The dynamic imagery rating ranged from 0 = “Did not trigger any moving images in me at all” to 100 = “Triggered a lot of moving images in me” (translated from German).

Participants were instructed at the beginning of each block that they should keep their eyes closed for the duration of each music excerpt to promote concentration and introspection and that they would be alerted as to when to close their eyes again after providing their ratings on a previous trial (‘So that you can always close your eyes in time, a visual countdown will appear before each piece of music, announcing the beginning of the respective piece of music’). Once ratings were provided (no time limit was given for providing ratings), a 4-s visual countdown (a short horizontal line that decreased in length with each second to become a dot at music onset) announced the start of the next listening trial. See Figure 3-1 for a summary of the trial procedure.

Figure 3-1

Procedure summary of each trial.



3.2.5. EEG recording

EEG was recorded using sintered Ag/AgCl active electrodes (suitable for reducing noise by amplifying the signal close to the source) and two 16-channel USB biosignal amplifiers (g.USBamp, g.tec medical engineering GmbH, Austria). A 530 Hz antialiasing filter was applied during recording, and the EEG recording sampling rate was set to 1200 Hz. A 15-channel 10–20 system cap was used, consisting of the following electrodes: AF3, AF4, F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, POz, PO7, Oz, and PO8. The reference channel was placed on the right mastoid, and the ground electrode was placed on the atlas (i.e., the top of the spine/back of the neck). For re-referencing offline to a non-lateralized reference, an additional electrode was placed on the left mastoid. Electrode impedance was maintained beneath 10 kΩ.

3.2.6. EEG data analyses

The EEG data was imported into MATLAB using EEGLab (Delorme & Makeig, 2004) functions and pre-processed using the FieldTrip toolbox (Oostenveld et al., 2011). The data was downsampled to a sampling rate of 200 Hz and filtered using a low-pass filter at 50 Hz. The data was subsequently segmented into epochs of 32 s, which comprised a 2-s pre-trial baseline phase and a 30-s main trial phase. All data was re-referenced to the average activity of the right and left mastoid channels.

The data (1,008 trials: 42 participants, 24 visual imagery main trials per participant, excluding practice trials) was visually assessed for artefacts. This allowed identification of channels that were faulty or that displayed extreme levels of variance. Across all participants, rejected channels (on average 1.7 and no more than four per participant) were interpolated with the average of neighbouring channels. Next, an independent component analysis (ICA) using the Runica algorithm (which implements the logistic infomax algorithm from EEGLab) was run. Spatial topographies were plotted, and individual components that were visually identified as eye movements, eye blinks, or localised electrode activity were noted and rejected from the data (on average 1.24 components per participant removed).

A Fast Fourier Transform frequency decomposition was carried out using a Hanning taper, with power computed in 1-s non-overlapping segments. Frequencies were extracted between 8 and 13 Hz for alpha band, between 14 and 30 Hz for beta band, and between 30 and 45 Hz for gamma band. For exploratory analyses, the alpha band was further subdivided into lower [8–10 Hz] and upper alpha [11–13 Hz]. The oscillatory power was baseline-corrected by subtracting the mean power of the 2-s pre-stimulus interval, separately for each trial within each channel. Specifically, for each trial within each channel, the average power in the prior two seconds (baseline period) was subtracted from power in each 1-s segment of the 30-s main trial.

3.2.7. Statistical analyses

My primary aim was to examine the relationship between visual imagery ratings (static and dynamic) and different forms of oscillatory activity (alpha, beta, and gamma power), and to determine how these relationships differed as a function of brain areas and time period. To this end, EEG channels were grouped into three regions of interest (ROI): Frontal (AF3, AF3, F3, Fz, and F4), Centro-Parietal (C3, Cz, C4, P3, Pz, and P4), and Parieto-Occipital (POz, PO7, Oz, and PO8), and the 30-s time windows of the main trials were divided into two phases: the First Half, comprising the first 15 s of a trial, and the Second Half, comprising the final 15 s of a trial. Here it is important to note that due to only having a single static and dynamic imagery rating for the whole trial (rather than continuous imagery ratings over time), my consideration of time effects are necessarily on a macro-level and are largely motivated by the finding that music-evoked visual imagery occurs on average around 12 s after music onset (Day and Thompson, 2019). Given that this estimation of imagery onset lies within the first half of the musical trials and given that dynamic imagery may nevertheless be expected to continue to change over time relative to static imagery, it seemed relevant to ask how patterns of activity in the first and second halves of the trials differed for the two types of imagery.

All statistical analyses were carried out using R (Version 4.2.3; R Core Team, 2018) and linear mixed models were estimated using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) R packages, the latter of which provided the t and p-values for the models. Given my main aim (to observe how static and dynamic ratings influenced oscillatory power differently as a function of brain areas and time period), I estimated, for each frequency band, a restricted maximum likelihood linear mixed model with oscillatory power as dependent variable, and static and dynamic imagery, and the interactions between each of these rating types with time period (First Half and Second Half) and ROI (Frontal, Centro-Parietal, and Parieto-Occipital), as fixed effects. Random effects were an intercept for participant and a nested random intercept between musical excerpt and excerpt type (joyful, neutral, fearful). Where this extra dimension in the random effects led to failed convergence in models, random effects were simplified by excluding the nested intercept and retaining only participant and excerpt type as random effects.

A Pearson's correlation coefficient showed static and dynamic imagery to correlate negatively but very weakly with each other, $r(1{,}006) = -0.094$, $p = 0.003$. I included both ratings within the same models to allow for the observation of the potentially different influences of each on the dependent variable.

See Table B.2 in Appendix B for a full summary of all omnibus models across all the frequency bands of interest. Follow-up models were run to explore any significant interactions emerging from the above models for each frequency band.

In an additional set of exploratory analyses, I analysed the influence of musical training on the pattern of EEG findings. To this end, musical training was dichotomised into high and low scores using a median split. Once again, I estimated for each frequency band a linear mixed model with oscillatory power as dependent variable, and as fixed effects: static and dynamic imagery, and interactions between these and musical training (High and Low) and time period. Random effects in each model were an intercept for participant and a nested random intercept between musical excerpt and excerpt type (joyful, neutral, fearful).

3.3. RESULTS

3.3.1. Analysis of alpha power

The overall model predicting alpha band showed main effects of static, $F(1, 435{,}460) = 4.89$, $p = 0.027$, and dynamic imagery, $F(1, 728) = 21.79$, $p < 0.001$, whereby high ratings were associated with suppression in alpha (see Table 3-1 for means and standard deviations of the two rating types), a main effect of time period, $F(1, 453{,}518) = 151.35$, $p < 0.001$, whereby there was less alpha power in the second compared to the first half of the trial, and also a main effect of ROI, $F(2, 453{,}518) = 295.10$, $p < 0.001$, whereby there was less alpha power in the frontal area, followed by the parieto-occipital area, then the centro-parietal area (see Table 3-2 and Figure 3-2 for means and standard deviations of oscillatory power for the three frequency bands across the two time periods and three regions of interest. Also see Figure B.1 in Appendix B for a frequency power plot displaying power across all trials and all frequency bands).

There were also interactions found between static imagery and time period, $F(1, 453{,}518) = 32.62$, $p < 0.001$, and between dynamic imagery and time period, $F(1, 453{,}518) = 38.28$, $p < 0.001$. To explore the significant interactions between static imagery and time period, a model was run to examine the relationship between static imagery ratings and alpha power for each time period separately. As also illustrated in Figure 3-3A and C, these revealed a significant negative association between alpha and static imagery (i.e., higher ratings in static imagery were associated with reduced alpha power) in the first half of the trial, ß = −0.00189, SE = 0.00134, $t(1.61) = -3.65$, $p < 0.001$, but a significant positive relationship in the second half, ß = 0.00328, SE = 0.00128, $t(2.10) = 2.57$, $p = 0.010$. Similarly, to explore the significant interaction between dynamic imagery and time period, a model was once more run for each time period separately. These revealed that dynamic imagery ratings' negative relationship with alpha power was significant in the first half, ß = −0.00550, SE = 0.00129, $t(1290.48) = -4.27$, $p < 0.001$, but non-significant in the second half, ß = −0.00040, SE = 0.00123, $t(2.56) = -0.32$, $p = 0.746$.

There was a significant interaction between ROI and both static, $F(2, 453{,}518) = 33.10$, $p < 0.001$, and dynamic imagery, $F(2, 453{,}518) = 121.88$, $p < 0.001$. To explore the significant interaction between static imagery and ROI, a model examining the relationship between power and ratings was run for each ROI separately. While no significant effects were found with respect to the frontal, ß = −0.00098, SE = 0.00080, $t(1.41) = -1.23$, $p = 0.218$, centro-parietal, ß = −0.00008, SE = 0.00161, $t(1.51) = -0.05$, $p = 0.961$, or parieto-occipital areas, ß = 0.00092, SE = 0.00209, $t(9.04) = -0.440$, $p = 0.660$, the interaction likely reflected a tendency for the relationship to be more systematically negative in frontal areas than in the other two ROIs (see also Figure 3-3B). Finally, following up the significant interaction between dynamic imagery and ROI, it was revealed that, as for static imagery, there was a tendency for the relationship to be more negative in frontal and posterior areas; specifically, while dynamic imagery ratings were significantly negatively linked to frontal alpha, ß = −0.00267, SE = 0.00076, $t(1.38) = -3.49$, $p < 0.001$, and parieto-occipital alpha, ß = −0.00441, SE = 0.00199, $t(956.51) = -2.22$, $p = 0.027$, there was no significant effect for the centro-parietal area, ß = −0.00106, SE = 0.00121, $t(1.81) = -0.87$, $p = 0.383$.

Finally, there was a significant interaction between ROI and time period, $F(2, 431{,}941) = 8.01$, $p < 0.001$. However, this was not explored further due to the current investigation's focus on visual imagery effects. No other main effects or interactions were significant.

Table 3-1

Summarising the descriptive statistics of static and dynamic visual imagery across the three types of musical excerpts presented to participants (joyful, neutral, and fearful).

	Mean Rating (Standard Deviation)			
	Joyful	Neutral	Fearful	Overall
Static Imagery	28.26 (22.98)	34.75 (29.25)	30.10 (22.65)	31.04 (25.29)
Dynamic Imagery	68.66 (24.91)	26.43 (24.92)	51.90 (27.85)	49.00 (31.21)

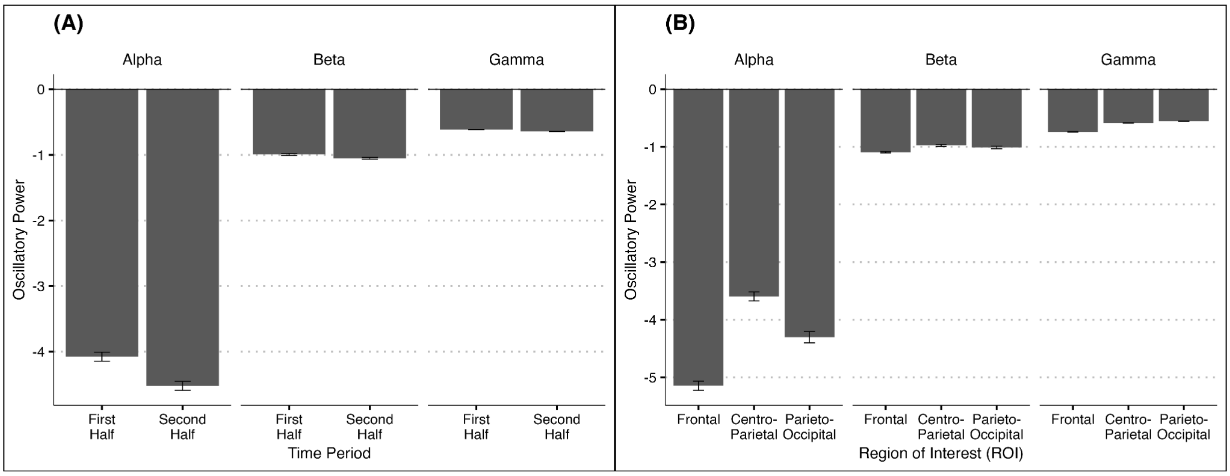
Table 3-2

Summarising the descriptive statistics of the frequency bands of interest across the two time periods (First Half and Second Half) and the three regions of interest (ROI; Frontal, Centro-Parietal, and Parieto-Occipital).

	Mean Power (Standard Deviation)				
	Time Period		**ROI**		
	First Half	Second Half	Frontal	Centro-Parietal	Parieto-Occipital
Alpha	–4.08 (32.84)	–4.52 (32.55)	–5.14 (30.85)	–3.59 (33.17)	–4.30 (34.15)
Beta	–0.99 (7.18)	–1.05 (6.64)	–1.09 (5.17)	–0.97 (7.25)	–1.01 (8.20)
Gamma	–0.62 (1.81)	–0.64 (2.16)	–0.74 (2.51)	–0.59 (1.92)	–0.55 (1.23)

Figure 3-2

Illustrating oscillatory power across the three frequency bands (alpha, beta, and gamma), including error bars depicting ± standard error of the mean. (A) Average power across the two time periods (First Half and Second Half). (B) Average power across the three regions of interest (ROI; Frontal, Centro-Parietal, and Parieto-Occipital).



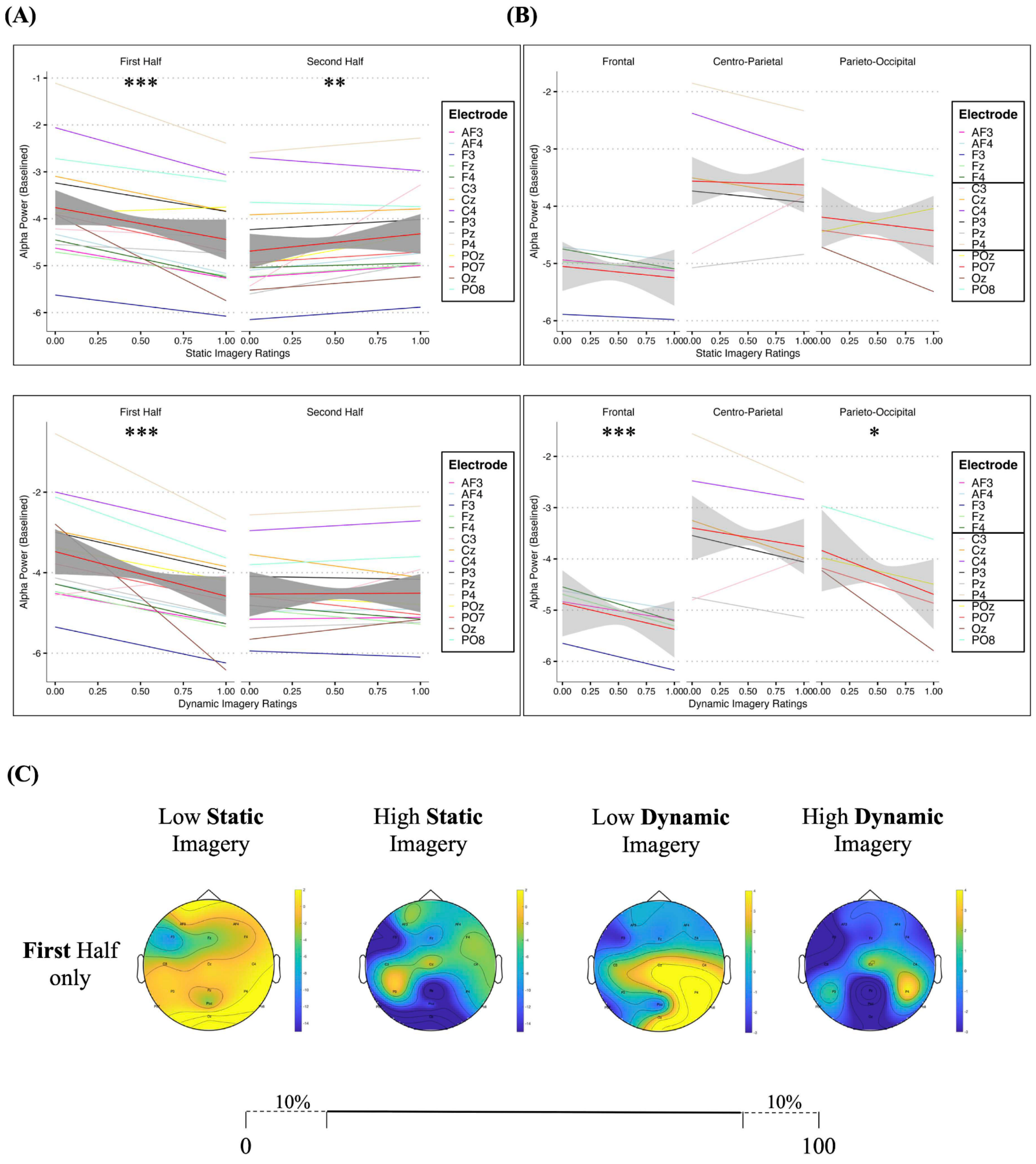


Figure 3-3

*Alpha power associated with static and dynamic imagery. Asterisks denote model significance levels: n.s. = non-significant, * p < 0.05, ** p < 0.01, *** p < 0.001. (A) Scatterplots showing alpha power as a function of static and dynamic imagery ratings, once the other had been regressed out, for two different time windows (First Half and Second Half). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (B) Scatterplots showing alpha power as a function of* *imagery ratings, once the other rating had been regressed out, for three regions of interest (Frontal: AF3, AF4, F3, Fz, and F4; Centro-Parietal: C3, Cz, C4, P3, Pz, and P4; and Parieto-Occipital: POz, PO7, Oz, and PO8). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (C) For illustrative purposes, I present topoplots showing patterns of alpha power (baselined to 2 s prior to the music) from trials associated with the upper 10% (denoted High) and lower 10% (denoted Low) of static and dynamic imagery ratings for the first half only.*

3.3.2. Analysis of beta power

The overall model for beta band showed a significant main effect of static imagery, $F(1, 453{,}488) = 176.64$, $p < 0.001$, and of dynamic imagery, $F(1, 281{,}826) = 9.50$, $p = 0.002$. High static imagery ratings were associated with higher beta power, whereas high dynamic imagery ratings were associated with lower beta power. There was also a main effect of time period, $F(1, 453{,}518) = 9.92$, $p = 0.002$, whereby there was higher beta power in the first half than in the second, as well as a main effect of ROI, $F(2, 453{,}518) = 22.69$, $p < 0.001$, showing there to be less beta power in the frontal area, followed by parieto-occipital, then the centro-parietal area.

There was also a significant interaction between static imagery and time period, $F(1, 453{,}518) = 90.52$, $p < 0.001$, and between dynamic imagery and time period, $F(1, 453{,}518) = 99.62$, $p < 0.001$. To explore the significant interactions between time period and both static and dynamic imagery, I ran four follow-up models. These showed a significant positive relationship between static ratings and beta power in the first time period, ß = 0.01193, SE = 0.00065, $t(2.26) = 18.39$, $p < 0.001$, but no significant relationship in the second, ß = 0.00085,

$SE = 0.00059$, $t(2.27) = 1.44$, $p = 0.150$. In contrast, dynamic imagery had a significant negative relationship with beta power in the first time period, ß = −0.00674, $SE = 0.00063$, $t(7.55) = -10.70$, $p < 0.001$, and a significant positive relationship in the second time period, ß = 0.00133, $SE = 0.00058$, $t(1.54) = 2.31$, $p = 0.021$. See Figure 3-4A and C which visualise beta power against dynamic imagery ratings across the two time periods.

Further, there was a significant interaction between static imagery and ROI, $F(2, 453{,}518) = 6.67$, $p = 0.001$. Exploration of the significant interaction between static imagery and ROI revealed significant positive relationships between static imagery and beta power in all the ROIs that nevertheless decreased in size from the front to the back of the head (frontal area, ß = 0.00642, $SE = 0.00048$, $t(1.51) = 13.31$, $p < 0.001$; centro-parietal area, ß = 0.00623, $SE = 0.00074$, $t(1.80) = 8.42$, $p < 0.001$; parieto-occipital area, ß = 0.00657, $SE = 0.00104$, $t(1.21) = 6.30$, $p < 0.001$).

There was further a significant interaction between dynamic imagery and ROI, $F(2, 453{,}518) = 7.17$, $p < 0.001$, as well as a significant three-way interaction between dynamic imagery, time period, and ROI, $F(2, 453{,}518) = 9.19$, $p < 0.001$. I summarise the latter interaction as it provides an extra dimension of detail. Exploration of this three-way interaction revealed a significant negative relationship between dynamic imagery and beta power in the frontal area in the first half, ß = −0.00612, $SE = 0.00052$, $t(3.91) = -11.74$, $p < 0.001$, but a non-significant negative relationship in the second half, $\beta = 0.00043$, $SE = 0.00078$, $t(4.40) = -0.55$, $p = 0.581$. Further, there was a significant negative relationship between dynamic imagery and beta power in the centro-parietal area in the first half, $\beta = -0.00494$, $SE = 0.00105$, $t(4.52) = -4.72$, $p < 0.001$, but a non-significant positive relationship in the second half, $\beta = 0.00152$, $SE = 0.00098$, $t(3.45) = 1.55$, $p = 0.120$. Finally, there was also a significant negative relationship between dynamic imagery and beta power in the parieto-occipital area in the first half, $\beta = -0.01031$, $SE = 0.00161$, $t(1.92) = -6.41$, $p < 0.001$, as well as significant positive relationship in the second half, $\beta = 0.00305$, SE = 0.00121, $t(3.63) = 2.52$, $p = 0.012$. No other main effects or interactions were significant.

Finally, there was a significant interaction between ROI and time period, $F(2, 453{,}518) = 3.59$, $p = 0.027$. However, again, this was not explored due to the current investigation's focus on visual imagery effects. No other main effects or interactions were significant. See also Figure 3-4B which visualises beta power against static and dynamic imagery ratings across regions of interest.

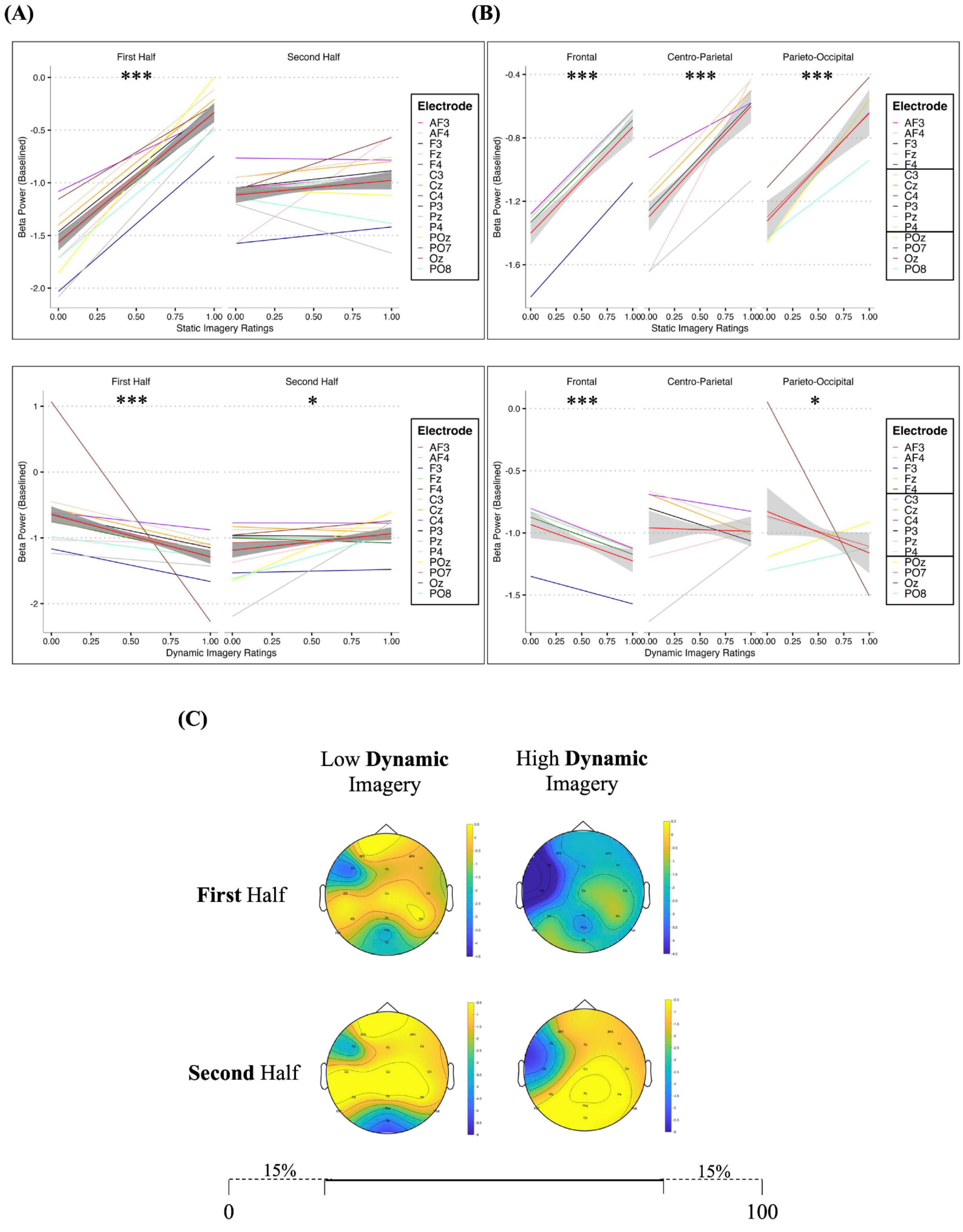


Figure 3-4

Beta power associated with static and dynamic imagery. Asterisks denote model significance levels: * $p < 0.05$, *** $p < 0.001$. *(A) Scatterplots showing beta power as a function of static and dynamic imagery ratings, once the other had been regressed out, for two different time windows (First Half and Second Half). For illustrative purposes, the x-axis is scaled to between*

0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (B) Scatterplots showing beta power as a function of imagery ratings, once the other rating had been regressed out, for three regions of interest (Frontal: AF3, AF4, F3, Fz, and F4; Centro-Parietal: C3, Cz, C4, P3, Pz, and P4; and Parieto-Occipital: POz, PO7, Oz, and PO8). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (C) For illustrative purposes, I present topoplots showing patterns of beta power (baselined to 2 s prior to the music) from trials associated with the upper 15% (denoted High) and lower 15% (denoted Low) of dynamic imagery ratings across the two different time windows.

3.3.3. Analysis of gamma power

The overall model predicting gamma band revealed a significant main effect of dynamic imagery, *F*(1, 14,847) = 5.73, *p* = 0.017, whereby an increase in dynamic imagery ratings was related to an enhancement of gamma power, as well as a significant main effect of ROI, *F*(2, 453,518) = 475.81, *p* < 0.001. There was also a significant interaction between dynamic imagery ratings and time period, *F*(1, 453,518) = 14.89, *p* < 0.001. To examine the interaction between dynamic imagery ratings and time period, a model was run for each time period separately. Dynamic imagery ratings showed a significant positive relationship with gamma power in the first time period, ß = 0.00038, SE = 0.00011, *t*(2.04) = 3.34, *p* < 0.001, but no effect in the second half, ß = −0.00007, SE = 0.00016, *t*(5.55) = −0.42, *p* = 0.673. See Figure 3-5A and C which visualise gamma power against static and dynamic imagery ratings across the two time periods.

There was a significant interaction between static imagery ratings and ROI, *F*(2, 453,518) = 222.20, *p* < 0.001, and between dynamic imagery and ROI, *F*(2, 453,518) = 24.45,

$p < 0.001$. Following up the interaction between static imagery and ROI revealed non-significant positive relationships between static imagery ratings and gamma in all three ROIs but indicated that these positive relationships were stronger in the centre than in the front and back of the head (frontal area, ß = 0.00011, SE = 0.00019, $t(1.41) = 0.59$, $p = 0.554$; centro-parietal area, ß = 0.00026, SE = 0.00015, $t(1.76) = 1.72$, $p = 0.085$; parieto-occipital area, ß = 0.00002, SE = 0.00012, $t(1.17) = 0.15$, $p = 0.883$). Finally, following up the interaction between dynamic imagery ratings and ROI revealed a significant relationship in the centro-parietal area only, showing that effects generally became more enhanced in the centre of the head than in the front and back (frontal area, ß = 0.00004, SE = 0.00018, $t(1.67) = 0.22$, $p = 0.826$; centro-parietal area, ß = 0.00036, SE = 0.00015, $t(5.35) = 2.44$, $p = 0.015$; parieto-occipital area, ß = 0.000002, SE = 0.00011, $t(3.10) = 0.02$, $p = 0.982$). See also Figure 3-5B which visualises gamma power against static and dynamic imagery ratings across regions of interest. No other main effects or interactions were significant.

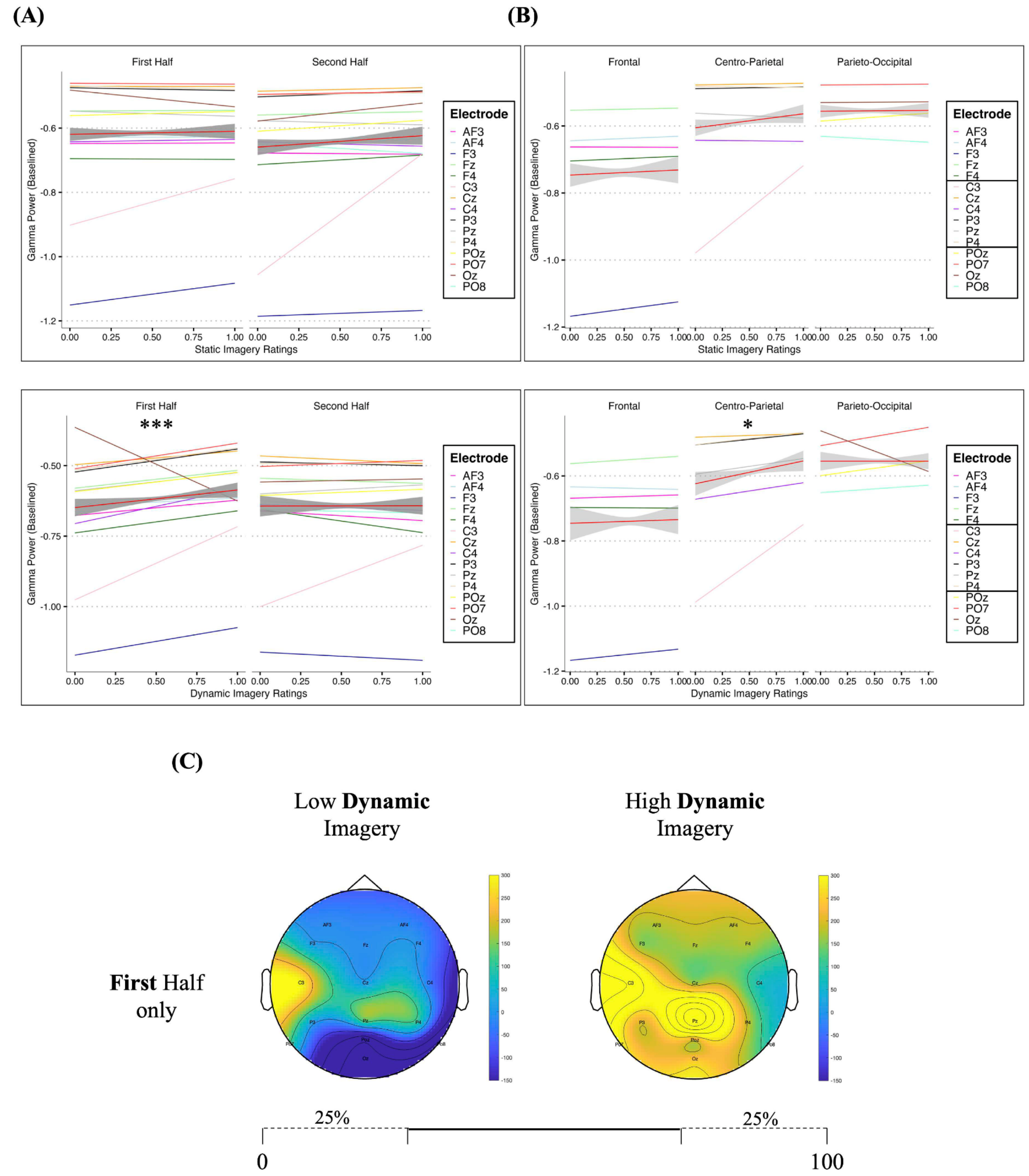


Figure 3-5

*Gamma power associated with static and dynamic imagery. Asterisks denote model significance levels: * p < 0.05, *** p < 0.001. (A) Scatterplots showing gamma power as a function of static and dynamic imagery ratings, once the other had been regressed out, for two different time windows (First Half and Second Half). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence* *interval. (B) Scatterplots showing gamma power as a function of imagery ratings, once the other rating had been regressed out, for three regions of interest (Frontal: AF3, AF4, F3, Fz, and F4; Centro-Parietal: C3, Cz, C4, P3, Pz, and P4; and Parieto-Occipital: POz, PO7, Oz, and PO8). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (C) For illustrative purposes, I present topoplots showing patterns of gamma power (baselined to 2 s prior to the music) from trials associated with the upper 25% (denoted High) and lower 25% (denoted Low) of dynamic imagery ratings for the first half only.*

3.3.4. Exploratory analyses: lower and upper alpha

Current evidence is mixed regarding a potentially nuanced role of upper and lower alpha in visual imagery: for example, while Petsche et al. (1996) linked visual imagination to greater power suppression in upper (than lower) alpha, Gualberto Cremades (2002) suggested that there is greater attenuation of lower (than upper) alpha particularly in those imagery tasks for which high attention and arousal is required.

In a set of exploratory analyses, I therefore examined how upper and lower alpha differ in their relationship to visual imagery by estimating the same modelling analysis (as carried out for alpha, beta, and gamma bands) for lower and upper alpha bands separately (see Appendix B for a detailed outline of the results, and see Table B.3 and Figure B.2 for means and standard deviations of lower and upper alpha power across the two time periods and three regions of interest).

Analysis of lower alpha showed a pattern of results that was almost identical to that seen for the full alpha band. Significant suppression in the low alpha frequencies was found more clearly in the first half of a trial than in the second half in response to both static and dynamic imagery. Additionally, trends towards suppression in the low alpha frequencies in all ROIs in response to high static and dynamic imagery were found, in addition to a significant suppression of frontal lower alpha power in response to dynamic imagery. In contrast, analysis of the upper alpha band, while largely similar to overall alpha, failed to show the strength of static imagery-related alpha suppression that was seen in lower alpha.

Further, while suppression in the frontal area associated with dynamic ratings was, similarly to the overall alpha band, significant in both lower and upper alpha, a frontal alpha suppression effect for static ratings – that was not seen in the overall or lower alpha model – was found when considering upper alpha alone.

In sum, the findings suggest that parieto-occipital lower alpha may be similarly involved in both visual imagery types but that differences between lower and upper alpha may reflect separate functionalities of static and dynamic imagery. See Figures B.3 and B.4 in Appendix B for visualisations of these results.

3.3.5. Exploratory analyses: effects of musicianship

There is little neuroscientific research concerning the relationship between musical training and visual imagery, though some behavioural investigations suggest that training enhances the vividness of music-evoked visual imagery (Küssner & Eerola, 2019), enables faster visual imagery formation through improved sensorimotor integration (Brochard et al., 2004), and leads to mental rehearsal of motor performance through kinaesthetic imagery (Lotze, 2013). EEG research shows that musical training improves the coactivation of auditory and sensorimotor processes in anterior brain (Bangert & Altenmüller, 2003; Klein et al., 2015; Trainor et al., 2009). In a final set of exploratory analyses, I therefore examined the influence that musical training scores may have on the patterns of neural activity that accompany visual imagery (see Appendix B for a detailed outline of the results).

In brief, I observed, with regard to alpha power and static imagery, that those with more musical training showed the suppression effect (in the first time window) to a greater extent than those with less training (although the rebound in the second was greater for those with less training). Interestingly, with regard to alpha and dynamic imagery, less trained participants showed the alpha suppression effect more in the first time window while more trained showed it to a greater extent in the second time window.

Further, while beta in response to static imagery did not show much difference between the two musicianship groups, the beta suppression effect in response to dynamic imagery (in the first time window) was greater in less trained than trained individuals. Finally, with regard to gamma, it was seen that the previously reported positive relationship between gamma and dynamic imagery was driven particularly by those with low training.

3.4. DISCUSSION

Behavioural studies have demonstrated music's propensity to elicit visual imagery (Balteş & Miu, 2014; Hashim et al., 2020; Juslin, 2013; Juslin & Västfjäll, 2008; Küssner & Eerola, 2019; Taruffi & Küssner, 2019, 2022) that varies in terms of dynamicity (Eitan & Granot, 2006; Johnson & Larson, 2003). The aim of the current study was to examine the neural signatures of music-evoked visual imagery and to determine the extent to which static and dynamic imagery can be seen reflected in differing patterns of neural oscillations.

Based on literature implicating the modulation of different frequency bands in the visual and motor brain regions in response to static and dynamic (De Lange et al., 2008; Fachner et al., 2019; Luft et al., 2019; Menicucci et al., 2020; Neuper & Pfurtscheller, 1996), I explored three oscillatory bands (alpha, beta, and gamma) in three main brain regions of interest (Frontal, Centro-Parietal, and Parieto-Occipital). Further, based on preliminary evidence regarding the timeframe within which music-evoked visual imagery can occur (Day & Thompson, 2019), I also considered how patterns of activity differ in the first (by which point visual imagery has likely occurred) and second halves of the thirty-second listening experience.

First, in line with the notion that visual imagery recruits similar brain areas and mechanisms to visual perception, I predicted that both static and dynamic visual imagery ratings would show a negative relationship with alpha as well as (particularly for dynamic imagery) a positive relationship with gamma in the parieto-occipital region (De Lange et al., 2008; Fachner et al., 2019; Lehmann et al., 2001; Luft et al., 2019; Sepúlveda et al., 2014). I further predicted that dynamic imagery, due to its recruitment of motor areas of the brain, would be associated with potentially complex patterns of beta desynchronisation and synchronisation (Menicucci et al., 2020; Neuper & Pfurtscheller, 1996; Salmelin et al., 1995; Villena-González et al., 2018; Zabielska-Mendyk et al., 2018).

3.4.1. Posterior alpha suppression during visual imagery

In line with past literature that has found strong links between visual imagery generation and occipital alpha activity (Cooper et al., 2003; Fachner et al., 2019; Sousa et al., 2017; Xie et al., 2020), I found evidence for parieto-occipital alpha suppression as a function of visual imagery, although this suppression effect with regard to time period more specifically was only present in the first half of the piece for both imagery types (i.e., within the 12 s period by which visual imagery is held to occur; Day & Thompson, 2019) and, for static imagery, even turned to an enhancement effect in the second half of the listening experience.

Here, I speculate that the ratings-related alpha suppression effect was limited to the earlier time window due to visual imagery having already emerged (or not) within this 12 s period (whether due to the affordances of the stimuli or even deliberate action on the part of the listener). Indeed, it is possible that this earlier period is what listeners base their imagery ratings on. In turn, I speculate that the consequent positive relationship between alpha power and static imagery ratings seen in the second half may reflect a rebound (increase) in alpha levels that is commensurate with the initial drop seen in the first half.

The visual imagery ratings and alpha relationships in the second half of the listening experience are interesting as they likely reflect the intrinsic difference between music-evoked static and dynamic visual imagery. Indeed, while I did not make this explicit prediction, it is possible that (in contrast to the tendency for alpha levels to show a rebound effect in the second half of the listening experience with respect to static imagery ratings) dynamic imagery ratings do not show an alpha rebound effect because dynamic imagery is constantly changing over the course of the listening experience and thus keeps alpha levels low (potentially at floor levels).

Here, it is, however, important to note that the low temporal resolution of the ratings prevents me from drawing detailed conclusions on the differences between static and dynamic imagery, and thus remains a key limitation of the current research. The current design, whereby ratings are provided only after each listening trial, fails to provide information on the dynamics of the prevalence and magnitude of visual imagery over the course of a music excerpt. Thus, my conclusions remain speculative, and further studies would greatly benefit from designs that enable the experience of visual imagery to be collected over time. Such designs would allow the neural correlates of static and dynamic imagery to be more carefully disentangled and would also throw light on how different types of imagery may be induced by different acoustic features.

3.4.2. Dynamic imagery associated with different patterns of activity from static imagery

I predicted that dynamic imagery would be similar to static imagery in terms of posterior alpha suppression but potentially more notably associated with gamma enhancement than static imagery. These predictions proved to be mostly accurate. In terms of posterior alpha suppression, I observed that dynamic imagery was associated with alpha suppression although unlike static imagery, dynamic imagery was not associated with a tendency to return to original alpha levels over the course of the imagery experience.

In terms of gamma activity, only dynamic imagery showed parieto-occipital gamma enhancement (De Lange et al., 2008). This difference found between dynamic and static imagery with respect to gamma oscillations is in line with past assertions that gamma oscillations are most evident when imagery is complex (De Lange et al., 2008; Sepúlveda et al., 2014), as can be assumed for more dynamic forms of imagery content.

Furthermore, with regard to additional differences between dynamic and static imagery, I predicted that dynamic imagery may be associated with both synchronisation and desynchronisation of beta frequency band in frontal, centro-parietal, and parieto-occipital areas. Indeed, here I showed that dynamic imagery was associated with a suppression (desynchronisation) in all regions of interest followed by an enhancement (synchronisation) of beta power in the parieto-occipital area specifically; this is in contrast to static imagery which was associated with beta enhancement in the first time period only and across all regions of interest. In line with previous results, I suggest that the beta suppression seen with dynamic imagery in the first 15-s time window reflects the recruitment of motor regions (Menicucci et al., 2020; Zabielska-Mendyk et al., 2018) and that the subsequent enhancement reflects the rebound in beta that has been reported to occur following such motor activity (Neuper & Pfurtscheller, 1996; Salmelin et al., 1995).

With regard to the unexpected beta enhancement seen as a function of static imagery ratings in the first time period, I propose – as has previously been suggested – that this may reflect the cross-modal processing of visual and auditory information simultaneously (Villena-González et al., 2018). As dynamic imagery is just as likely to involve such cross-modal processing, it is important to put forward the possibility that – along with any beta power rebound effect (Neuper & Pfurtscheller, 1996) – such cross-modal processing also drives the positive relationship between beta and dynamic imagery that is seen in the second time period.

Finally, it is important to note that unlike static imagery, dynamic imagery was associated with robust frontal alpha suppression. Previous work has shown dynamic imagery to be associated with greater frontal alpha power when compared to static imagery (Sousa et al., 2017); and others have suggested that this may be due to the higher internal processing demands and task complexity (Cooper et al., 2003; Klimesch et al., 2007; Sauseng et al., 2005; Schomer & Silva, 2012) of dynamic imagery compared to static imagery. Even though my findings seem opposed to those from Sousa et al. (2017), it is important to note that a direct comparison with their work is difficult since their study compared two different conditions (static and dynamic), whereas I measured static and dynamic imagery using subjective ratings for each. Further work is needed to corroborate my finding of frontal alpha suppression in response to dynamic content of visual imagery.

An additional exploration involved examining potential differences between lower and upper alpha in their associations with static and dynamic imagery. The results demonstrated minimal differences between overall alpha power and lower alpha power with regard to power suppression in response to static and dynamic imagery, suggesting that the lower band may be an effective reflection of both types of visual imagery content. The findings nevertheless also show that certain effects may be present in upper but not lower alpha (e.g., frontal alpha suppression in response to static ratings) thus suggesting that these sub-bands may indeed play distinct roles. Taken together, my findings support past literature proposing distinct functions of lower and upper alpha power (Petsche, 1996).

The final exploration involved examining the influence of musical training on the relationship between oscillatory power and visual imagery ratings. Though findings in previous research have been mixed (Aleman et al., 2000; Küssner & Eerola, 2019), patterns of results suggest that musical training influenced patterns of neural activity in nuanced ways. The differences in patterns of oscillatory activity – during visual imagery – that one should expect when considering how the musically trained brain differs from the typical brain is not yet well documented. Further studies are therefore needed to understand this potential influence of musical training on neural substrates of music-evoked visual imagery.

In summary, the results present a few unexpected differences between static and dynamic imagery with respect to alpha modulation but are in line with my predictions that i) both imagery types would drive posterior alpha suppression, ii) the two may show different effects over time (due to dynamic but not static imagery being associated with continuous changing imagery content), and iii) dynamic imagery, to a greater extent than static imagery, would show complex modulation of areas involved in motion processing (Sirigu & Duhamel, 2001; Thompson et al., 2009; Zacks, 2008).

3.4.3. Implications of the research

I provide preliminary evidence to show that the neural correlates of music-evoked visual imagery may include both posterior alpha suppression, and beta and gamma modulation. This research adds to similar previous research by Fachner et al. (2019) who also showed evidence of posterior alpha suppression but did not investigate gamma power. As such, my study highlights the fact that music is a powerful tool to study the neural correlates of visual imagery. Further, the consistency of my findings with the previous body of work on visual motor imagery can be taken as support for the idea that motion plays a significant role in music-evoked visual imagery (Antović et al., 2024; Dahl et al., 2022) and, as such, in the music listening experience more generally (Eitan & Granot, 2006; Johnson & Larson, 2003; Schaefer, 2022).

However, perhaps the most important implication of my research relates to future investigations of the neural correlates of dynamic visual imagery. Indeed, this literature, much of which has focused on athletes (Wilson et al., 2016) or used small samples to explore brain-computer interfaces (Sousa et al., 2017), offers mixed results as to the directionality and cortical localisation of effects related to the experience of dynamic or motion-related imagery. My findings, which are in line with the results of a meta-analysis suggesting a great overlap between visual motor imagery and kinaesthetic motor imagery (Filgueiras et al., 2018), suggest that music-evoked visual imagery could be particularly useful for future studies seeking to explore dynamic content of visual imagery.

That said, it is important to consider the possibility that studies using musical stimuli show neural effects that not only reflect visual imagery content but also features of the music that induce said imagery content. Here, I sought to account for the effects of music heard (on the differing brain signatures I reported for static and dynamic imagery) by including the musical pieces as random effects within my mixed modelling analysis approach. However, preliminary findings of the effects that musical features can have on visual imagery (Dahl et al., 2022; Herff et al., 2022; Juslin, 2019) suggest that future studies may want to control for such variables even more carefully.

Relatedly, past behavioural studies have reported high prevalence rates of visual imagery during music (Dahl et al., 2022; Küssner & Eerola, 2019; Vuoskoski & Eerola, 2015), leading me to view it as a useful reliable stimulus. However, a relevant concern is that insights from studying music may not all be generalisable to other instances of static and dynamic imagery. Here, I suggest that while some neural patterns may relate specifically to music's acoustic properties, this is the case for all other stimulus types that may be used for inducing imagery. Further, even though I pointed out that music-evoked visual imagery has been related to music's aesthetic appeal (Belfi, 2019), differing levels of aesthetic evaluation to differing imagery content is something that would be present in other imagery induction paradigms, including simpler paradigms where people are asked to imagine different scenes in the absence of any inducer. Future studies where such additional factors are a potential concern could seek to track such variables and account for them in the analysis approach.

As previously mentioned, future investigations should further seek to corroborate and extend my results by using alternative experimental designs that enable explicit comparison between music-evoked static and dynamic visual imagery. While the current study touches upon partially distinct processes that may underpin these two types of visual imagery, such further studies will be valuable.

3.4.4. Conclusion

The aim of Study 2 was to further our understanding of the oscillatory characteristics underlying two types of visual imagery content during music listening: static and dynamic. In line with my predictions, the findings corroborated past literature on the oscillatory signatures of visual imagery and revealed nuanced differences in signatures of static and dynamic content of visual imagery. Investigations into what listeners tend to imagine whilst listening to music is gaining traction. This study opens further avenues into the operationalisation of commonly occurring forms of visual imagery and into the examination of how they can be observed in neural data. The following chapter uses EEG once again to examine two further types of music-evoked visual imagery that pertain to its intentionality, spontaneous and deliberate visual imagery.

CHAPTER 4. EXPLORING THE NEURAL CORRELATES UNDERPINNING TWO INTENTIONALITY STATES OF MUSIC-EVOKED VISUAL IMAGERY

In Study 3, I combined the probe-caught experience sampling methodology with a 32-channel electroencephalography (EEG) in order to investigate the oscillatory characteristics of spontaneous and deliberate music-evoked visual imagery. Thirty participants listened with closed eyes to four blocks of music differing in familiarity and each of which spanned a range of genres. In response to regular thought probes sent throughout listening, participants indicated whether or not they had been experiencing visual imagery and, if they had, whether the experienced imagery was spontaneous or deliberate. Cluster permutation analyses on the time-frequency decomposed EEG data revealed alpha power suppression during visual imagery that was greater during spontaneous than deliberate imagery. While theta and delta bands did not discriminate either the experience of visual imagery or its intentionality subtypes, gamma power suppression was observed during visual imagery. The results thus replicated Study 2 and other prior findings of a role of alpha suppression in posterior areas during visual imagery. Further, it raised the possibility that more deliberate forms of music-evoked visual imagery may recruit visual processing areas to a lesser extent than spontaneous forms.

4.1. INTRODUCTION

The internalisation of one’s attention is a common mental action that leads to a variety of states including visual mental imagery (Pearson, 2019), mind wandering (Christoff et al., 2016; Martarelli et al., 2016), and creative thinking (Belfi et al., 2017; Luft et al., 2019). Music listening, an activity often used to escape the outside world (Taruffi et al., 2017), has a widely recognised ability to evoke visual imagery in listeners, with some studies suggesting music-evoked visual imagery occurs in at least 70% of listeners (Küssner & Eerola, 2019; Vuoskoski

& Eerola, 2015). Recent research is beginning to characterise the different forms that music-evoked visual imagery can take (e.g., Study 1 in Chapter 2; Dahl et al., 2022; Groves et al., 2023; Küssner & Eerola, 2019), and to reveal how these different forms are reflected in brain activity (e.g., Study 2 in Chapter 3). However, few studies into music-evoked visual imagery have specifically addressed the question of how often such experiences happen spontaneously and how often they are deliberately sought. Relatedly, no study has examined the extent to which spontaneous and deliberate forms of music-evoked visual imagery implicate differing patterns of neural activity.

In mind wandering research, authors have proposed the existence of two attentional modes of mind wandering, namely intentional mind wandering ("tuning-out") and unintentional mind wandering ("zoning-out") (Polychroni et al., 2022; Seli et al., 2016). While the intentionality of visual imagery has not yet been a focus of research, the literature nevertheless suggests that there is a basis for characterising visual imagery as being able to occur either spontaneously or with deliberate intention (Taruffi & Küssner, 2019). Specifically, alongside studies that have reported on deliberate forms of imagery following explicit instruction (Study 1 in Chapter 2; Cespedes-Guevara & Dibben, 2022; Day & Thompson, 2019; Hashim et al., 2020; Herff et al., 2022; Küssner & Eerola, 2019), other studies have examined and reported on more spontaneous forms of visual imagery, as can be evoked, for example, during deep meditation (Luft et al., 2019).

4.1.1. Capturing the occurrence of music-evoked visual imagery

Study 2 (in Chapter 3) of this thesis, where participants were asked to rate the extent of their static and dynamic imagery after music listening while EEG was recorded, was important in confirming that the qualities of music-evoked visual imagery are distinguishable in neural data. However, by requiring participants to report on how much imagery they experienced on each trial, that study arguably encouraged deliberate, rather than more spontaneous forms of imagery. In any case, by not directly inquiring as to whether any visual imagery that was experienced was spontaneously or deliberately generated, Study 2 and other previous studies have been unable to speak to how these intentionality states of visual imagery differ.

Recent mind wandering research has showcased the usefulness of the probe-caught methodology (Polychroni et al., 2022; Taruffi et al., 2017) for capturing nuanced differences in those moments when individuals shift their attention inward. One study on mind wandering during music listening by Taruffi and colleagues (2017) used a simple version of this method (whereby listeners were only probed once at the end of each excerpt) to examine the occurrence and nature of listeners' mind wandering experiences. The authors found that, compared to the experience of words, visual imagery was a prevalent aspect of participants' mind wandering experiences during music listening. This study was interesting in not just conceptualising visual imagery as a form of mind wandering but also showing that visual imagery has a very high occurrence rate during music listening. Critically, other studies from the authors have taken a similar approach and have provided useful findings regarding the incidence of music-evoked visual imagery in everyday settings with respect to intentionality and meta-awareness. For instance, it has been reported that 19% of mind wandering cases comprise visual imagery, and that 65% of music-related (as well as 59% of non-music-related) mind wandering occur spontaneously (Taruffi, 2021). In any case, given the simplicity and conciseness of the method and the effectiveness with which it allows measurement of both the occurrence (Weinstein, 2018) and neural correlates (Polychroni et al., 2022) of mind wandering states, the probe-caught method presents as an invaluable way to examine the incidence and neural correlates of visual imagery experiences of differing intentionality.

4.1.2. Neural correlates of music-evoked visual imagery

As previously discussed, a growing body of work shows that while seemingly highly subjective, the occurrence and the contents of visual imagery can nevertheless be seen reflected in neural activity (Cooper et al., 2003; Dijkstra, Bosch, et al., 2017; Dijkstra et al., 2019; Drever, 1955; Klimesch et al., 2007; Williamson et al., 1997; Xie et al., 2020). Study 2 (in Chapter 3) of this thesis confirmed that music-evoked visual imagery is associated with suppressed alpha power [8-13 Hz], and in doing so, corroborated findings from the only other study to examine music-related visual imagery (Fachner et al., 2019). However, in addition to attenuated alpha power in primarily posterior regions of the brain, research has also implicated other potential correlates of music-evoked visual imagery. These include gamma power in the occipital brain area associated with spontaneous imagery conjured during deep meditation (Luft et al., 2019), in addition to theta and delta bands in frontal, central, and parietal areas (Polychroni et al., 2022).

Study 2 showed that while both static and dynamic imagery led to alpha suppression during visual imagery formation as expected (Klimesch, 1999; Klimesch et al., 2007; Sousa et al., 2017; Xie et al., 2020), dynamic imagery was further associated with brain areas and oscillatory bands typically linked with motoric action and higher internal processing demands (De Lange et al., 2008; Menicucci et al., 2020; Sepúlveda et al., 2014; Zabielska-Mendyk et al., 2018). It is thus likely that alpha and other neural oscillations will also differ as a function of whether one is experiencing visual imagery spontaneously or due to deliberate action. With regard to alpha suppression effects, one possibility is that deliberate imagery will implicate less alpha suppression in visual areas as result of being driven by top-down mechanisms rather than by automatic lower-level visual processing (for a review, see Pearson, 2019). Specifically, it is possible that – unlike for spontaneous imagery, where images occur, automatically, freely and unbidden, and therefore implicate low-level visual processing areas – reports of deliberate forms of imagery will include situations in which any imagery that is successfully bidden is at least partly conceptual (i.e., less pictorial and potentially less vivid), and thus not as associated with visual processing areas as spontaneous imagery (Gilbert & Li, 2013; Mechelli et al., 2004).

Whether effects should be expected outside the alpha band is unclear. Findings of higher theta power during unaware mind wandering states have been linked to similar findings during meditative and absorption states (Polychroni et al., 2022). Similarly, delta power increases found during unaware mind wandering have been interpreted as indexing the maintenance of internal trains of thought through the inhibition of external interference (Polychroni et al., 2022). Given that visual imagery may be considered a form of mind wandering, one possibility is that, as for unaware mind wandering, spontaneous visual imagery will be associated with enhancements of both delta and theta activity. However, since intentionality of visual imagery (spontaneous experiences vs. deliberate experiences) is only partially related to the metacognitive awareness of imagery (i.e., spontaneous imagery, due to the lesser cognitive control involved, may be expected to be associated with lesser awareness than deliberate imagery, but can still reach conscious awareness), expecting similar patterns of activity as in the study from Polychroni and colleagues may not be wholly justified.

Taken together, research has provided evidence that music-evoked visual imagery has reliable neural correlates. However, still unclear is if and how the neural correlates of spontaneous forms of visual imagery during music listening and in general, may differ from neural correlates of more deliberate imagery. Such insights would improve understanding of how the brain mediates mental actions differing in their intentionality.

4.1.3. The present study

The current experiment sought to examine the neural correlates of deliberately and spontaneously occurring visual mental imagery in response to music during concurrent EEG recording. Using an ecologically valid probe-caught method (Polychroni et al., 2022; Taruffi et al., 2017; Weinstein, 2018), participants were regularly probed on both visual imagery occurrence and on the intentionality of any imagery throughout extended music listening trials. Specifically, using dichotomous responses, they were requested to report on whether they had or had not experienced any visual imagery immediately prior to receiving a probe (providing listeners with the option to report no visual imagery at all), and, if answered in the affirmative, were asked to indicate whether the formation of this imagery had been spontaneous or deliberate. Musical stimuli differing in terms of relaxation potential (relaxing vs. non-relaxing) as well as familiarity (experimenter-selected vs. participant-selected) were used, to allow generalisation of findings to the wide range of stimulus types that are typically experienced in everyday life (Jakubowski & Francini, 2022; Pereira et al., 2011). Using a variety of stimulus types also allowed investigation of how these factors influence the relative incidence of spontaneous, versus more deliberate forms of visual imagery. Finally, individual differences in musical training, general imagery ability and music preference were also measured to allow examination of if and how these seemed to influence imagery experiences.

In light of the presented research, I hypothesised the following:

1) All music-evoked visual imagery would be characterised by alpha power suppression in the posterior regions of the brain (Fachner et al., 2019; Sousa et al., 2017; Xie et al., 2020),

2) Deliberate visual imagery would be less associated with such alpha power suppression (than spontaneous imagery) due to its being a more top-down process and, as such, a less bottom-up vivid visual experience-led phenomenon,
3) Spontaneous visual imagery may be associated to some extent with increases in both theta and delta power; this, due to the fact that spontaneous imagery may be less likely to reach conscious awareness and may thus be more similar (than deliberate visual imagery) to unaware mind wandering (Polychroni et al., 2022),
4) Familiar music would result in greater incidence of music-evoked visual imagery experience (due to its greater association with autobiographical memory), as would higher overall general visual imagery ability and greater preference for the genres of the musical stimuli in question.

4.2. METHODS

4.2.1. Participants

Data was collected from 30 participants (aged 18-49, 20 female, 9 male, 1 non-binary; M = 27.03, SD = 7.5) who were recruited using poster advertisements, word-of-mouth, and a student credit scheme. An online survey was used to assess eligibility for the current research. The survey asked participants to report (on a 1-7 Likert scale, from 1 = 'Not at all' to 7 = 'Experience very strongly') the extent to which they experienced the following phenomena in response to music: chills or goosebumps, visual imagery, felt emotions, and personal memories (a selection of experiences were asked in order to help mask the intentions of the study). The primary inclusion criterion was that individuals report experiencing at least mild levels (providing a rating of 2 or upwards) of visual imagery in response to music. Those who reported never experiencing music-evoked visual imagery (i.e., provided a rating of 1) were considered ineligible and were thanked for their time.

4.2.2. Ethics statement

This research was approved by the Research Ethics Committee of the Department of Psychology at Goldsmiths, University of London. All participants provided online and in-person written informed consent before taking part. Eligible participants were provided with monetary compensation of £10 per hour for their time.

4.2.3. Obtaining self-selected stimuli

In addition to screening participants, the initial online survey asked eligible participants to provide examples of musical pieces that they tended to listen to in order to relax as well as up to four examples of pieces that they would *not* listen to when looking to relax (e.g., to heighten arousal instead). These questions were modelled after those used by Baltazar et al. (2020). Specifically, the first question asked: '*Imagine you are feeling anxious, stressed or nervous, but you need to calm down in order to be able to focus on your work. Whilst in this situation, you decide to listen to some music to help you relax. What would be some good examples of music pieces that would work for you in this kind of situation?*'. The second question asked: '*Please also think of music pieces you are familiar with and you like, but would not work well in this stressful situation. In other words, music that would not necessarily help you feel more relaxed.*' Participants were encouraged to provide 2-4 examples for each question and were asked to provide examples not containing lyrics or (where this was not possible) to provide acoustic alternatives to the lyrical versions.

4.2.4. Pilot survey for the selection of unfamiliar music of high and low relaxation potential

A pilot survey was conducted to obtain experimenter-selected musical excerpts for the main experiment (the full results of this pilot survey are provided in Appendix C). I recruited a total of 17 participants (aged 18 – 42, female = 10, male = 7, M = 26.71, SD = 7.04), and the survey included excerpts from electronic, classical, and jazz genres. The tracks were selected from a range of sources, but mostly from Marti-Marca et al. (2020), allowing me to vary pieces on the basis of valence and arousal as basic markers of the relaxation potential of a given track.

In total, twelve pieces were used in the pilot survey (see Table C.1. in Appendix C for the full list): six that were considered high in relaxation potential and six low in relaxation potential with the inclusion of two options from each genre (EDM, classical, and jazz) within each group. Participants rated (using a 1-7 Likert scale) parameters similar to those used by Baltazar & Västfjäll (2020), namely: energy (how intense or active the music was), acousticness (how much the track consisted of acoustic instruments or electronic sounds), valence (positivity/negativity of the music), danceability (suitability for dancing), and loudness (overall volume conveyed). Participants were also asked how relaxing they found the music, how much they liked it, and how familiar they were with it.

The final stimuli set used in the main experiment is indicated in Table C.1 in Appendix C in further detail. To be labelled as having high relaxation potential, tracks needed to be rated on a 1-7 Likert scale as having low (ratings of 3 or below) energy, low danceability, low loudness, and high (ratings of 4 and above) relaxingness. In contrast, low relaxation potential tracks needed to be rated with high energy, high danceability, high loudness, and low relaxingness. Acousticness was used to confirm that an equal number of acoustic and electronic tracks were included in the two groups. Valence and liking were collected to ensure that tracks were rated as being at least neutral or higher in terms of positivity and enjoyment. Familiarity ratings were obtained to ensure that tracks selected for the two groups were similarly unfamiliar (presented as '*How familiar are you with this excerpt?*' and rated along a 4-point Likert scale from 1 = 'As far as I know, I've never heard this excerpt before' to 4 = 'I know this excerpt, it is called/by: [*open-text response*]'). As observed in Table C.2 in Appendix C, familiarity values were generally low across all music tracks, confirming their unfamiliarity.

4.2.5. Materials and stimuli

Listening conditions. The listening task was divided into four conditions: an unfamiliar (i.e., experimenter-selected) relaxing, an unfamiliar non-relaxing, a familiar (i.e., participant-selected) relaxing, and a familiar non-relaxing music condition. In the event that the familiar music stimuli provided by participants in the screening survey did not together extend to 20 mins duration (the full length of each listening trial), excerpts were repeated, starting from the longest track to the shortest, until the full duration was reached.

Probe-caught measure. Visual imagery experience was assessed by presenting thought probes at pseudorandom intervals throughout the listening trials. Probes were presented every 40 s with a uniform ± 10-s jitter (i.e., all possible durations between 30 and 50 s across every listening condition). Phrasing and implementation of the thought probes followed previous mind wandering literature (Polychroni et al., 2022; Taruffi et al., 2017; Weinstein, 2018). The first probe prompted the participant to indicate if they were experiencing any visual imagery: '*Just before the probe, were there any images in your mind's eye?*', presented with the dichotomous response boxes: '*Yes*' and '*No*'. If the participants responded with '*Yes*', they were asked a follow-up question: '*Were your thoughts deliberate and actively generated or spontaneous and free-flowing?*', presented with the dichotomous response boxes: '*Deliberate*' and '*Spontaneous*'. Participants were given up to 4 s to answer each probe. These parameters were set to allow presentation of as many probes as possible (a total of up to 25 probes per listening condition and up to 100 probes over the four conditions), while allowing sufficient time in between probe presentations for the participant to return to the listening task.

The limited time given to respond was also to discourage participants from overthinking their answers and to avoid interrupting the listening experience for too long.

Affect ratings. At the end of each listening condition, I collected affect ratings on a visual analogue 1-7 Likert scale (‘*Please rate the following indices to best reflect your current state*’) with questions similar to those used by Baltazar et al. (2019): valence (negative–positive), energy (drowsy–alert), and tension (relaxed–tense). Additional thirteen questions further inquiring into the content of visual imagery experienced as well as general impressions regarding the music heard during the listening trials were also presented at the end of each block (see Table C.3 in Appendix C for a list of these additional questions). However, this data will be analysed and discussed elsewhere.

Individual differences. The 7-item Musical Training subscale of the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) was presented to gauge participants’ levels of prior musical training (e.g., ‘I have never been complimented for my talents as a musical performer’, ‘I am able to judge whether someone is a good singer or not’), using a 7-point Likert scale from 1 = ‘Completely Disagree’ to 7 = ‘Completely Agree’.

I further presented the 5-item Vision and 5-item Emotion subscales of the Plymouth Sensory Imagery Questionnaire (Psi-Q; Andrade et al., 2014), which assesses one’s general ability to experience mental imagery across multiple modalities (such as visual, tactile, emotion, olfactory, etc.; e.g., ‘Imagine the appearance of … a sunset’, ‘Imagine feeling … excited’). These items were rated on an 11-point Likert scale from 0 = ‘no image at all’ to 10 = ‘imagery as clear and vivid as real life’.

Finally, a selection of genres (classical, electronic dance music, jazz, and ambient) were presented to examine participants’ preferences for the musical genres presented within the experimenter-selected tracks (‘*Please indicate your basic preference for each of the following* *genres using the scale provided.*'). Participants rated their preference for each genre using a 7-point Likert scale from 1 = 'Dislike Strongly to 7 = 'Like Strongly'.

4.2.6. Procedure

The experiment was programmed and presented using the experiment-building software, OpenSesame (Version 3.3.8, Python 3.7.9; Mathôt et al., 2012). Participants sat in a dimly lit room and were presented with verbal as well as written instructions. The stimuli were played through headphones and participants were advised to adjust the volume to a comfortable level while listening to a sample listening track and to maintain that volume for the duration of the experiment. Participants completed – in randomised order – the four listening blocks with music differing in familiarity and relaxation potential, and with each block lasting approximately 25 mins. The full duration of the experiment was between 2 and 3 hours (including EEG preparation, verbal and written instructions, and potential questions). Prior to the start of the experiment, deliberate visual imagery was defined as *visual thoughts that are purposefully thought out*, whereas spontaneous visual imagery was defined as *appearing without any particular intention or control.*

For each of the four listening blocks, participants first stared at a fixation dot for 30 s, before being asked to provide ratings (using three affect ratings) on how they were feeling at that moment and then being presented with the 20-minute music block. Participants were informed that they should listen to the music presented, by immersing themselves with eyes closed, and that they would hear a brief clicking sound any time a probe question appeared on the screen. Participants were told they should open their eyes at that point, provide their answer, and then continue to listen with eyes closed. At the end of each block (indicated by a final click sound), participants were once more required to fixate for 30 s and to complete the affect questions, before answering additional questions about their visual imagery experience. To finish, participants filled in the Musical Training subscale of the Gold-MSI, the Vision and Emotion subscales of the Psi-Q, and the short musical genre preference survey.

4.2.7. EEG recording and analyses

EEG was recorded using the mobile Waveguard 32-channel cap from ANT Neuro, The Netherlands. The 32 channels used were: Fp1, Fpz, Fp2, AFz, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, M1, CP5, CP1, CP2, CP6, M2, P7, P3, Pz, P4, P8, POz, O1, Oz, O2. The cap was placed in accordance with ANT Neuro's 5% electrode placement system, a derivative of the 10-20 international system. Signals were amplified using the eego EEG recording amplifier, with the reference set to electrode CPz. Impedance levels were never above 20 kΩ. The EEG recording was set to a sampling rate of 500 Hz for all participants.

The EEG data was imported into MATLAB using EEGLab (Delorme & Makeig, 2004), and pre-processed and further analysed using functions from the FieldTrip toolbox (Oostenveld et al., 2011). Data segmentation, pre-processing and analysis choices were broadly in line with previous studies (Compton et al., 2019; Liu et al., 2021; Polychroni et al., 2022). The continuous data was first segmented into 34.7-s epochs: –12 s before the probe onset to 22.7 seconds after probe onset. The time window of interest was from –10 to 0 s relative to probe onset. However, the longer epoch was taken to allow a 700-ms baseline window within the interstimulus interval after probe onset and the start of the next probe trial, as well as to avoid edge artefacts.

The segmented data was resampled to a rate of 200 Hz. Further, a low-pass band filter was set at a frequency of 50 Hz and a notch filter at a frequency of 48-52 Hz. Bad or faulty electrodes were visually searched for by observing the raw data, but as none were found, no electrodes were removed or interpolated in this step. Artefacts were then visually assessed by scrutinising the time courses of each trial's raw data and replacing bad segments with NaN's (average seconds removed = 1.31, max = 6.32, min = 0.08). Next, independent component analysis (ICA) was used to identify and remove artefacts caused by eye movements, eye blinks, heartbeats, and further channel noise (average number of components removed = 4.7). Finally, trial ✕ channel spectra were visually assessed to reject whole trials that still exhibited high levels of variance even after prior artefact removal efforts (average number of trials excluded = 0.93). All the data was then re-referenced to the average of all the electrodes.

Finally, in preparation for statistical analyses, epoched data was subdivided into four groups corresponding to the four possible visual imagery states. These groups were *no visual imagery*, *all visual imagery*, *deliberate* visual imagery only, and *spontaneous* visual imagery only.

4.2.8. Time-frequency decomposition

A time-frequency decomposition of epochs was computed using a Hanning taper moving with 50 ms steps along the 34.7-s epoch for frequencies from 1 to 50 Hz. This decomposition was carried out using a frequency-dependent window length, with 7 cycles per time window. Spectral power for every trial was averaged then normalised by dividing it by the baseline level (i.e., the gain model; Grandchamp & Delorme, 2011).

A 700-ms baseline was selected as 15 to 15.7 s post probe onset, as it was late enough in the inter-stimulus interval to be sure that participants had closed their eyes again after responding to a probe (participants had up to 8 seconds to respond to the two questions) but early enough to avoid that substantial visual imagery had been formed again. Importantly, comparing this choice of baseline period against one that was earlier (10 s after probe onset) and another that was later (20 s after probe onset) showed that the nature of effects reported below (see section *4.4.1*) did not differ across the three baseline windows.

4.3. STATISTICAL ANALYSES

4.3.1. Spectral analysis: differences between visual imagery manifestations

Differences in the spectral power between the different visual imagery states (visual imagery vs. no visual imagery, spontaneous vs. deliberate, spontaneous vs. no visual imagery, and deliberate vs. no visual imagery) were analysed using non-parametric cluster-based paired permutation tests (Maris & Oostenveld, 2007), run for each of the oscillatory frequency bands of interest (delta [2-3 Hz], theta [4-7 Hz], alpha [8-13 Hz] and gamma [30-45 Hz] for completion) and using functions from the MATLAB-based Fieldtrip toolbox (Oostenveld et al., 2011).

These analyses were first carried out by calculating the observed test statistic by means of paired samples t-tests between the specified contrasts at each sample of the 3-dimensional data (channel vs. frequency vs. time). Those samples for which the t-statistic fell beneath the $\boldsymbol{\alpha} < 0.05$ threshold were selected and clustered in connected sets on the basis of temporal vs. spatial vs. spectral adjacency. The t-statistics within every cluster were then summed to calculate the cluster-level statistics. The maximum of these statistics served as the test statistics with which differences between visual imagery and manifestation states were determined.

Subsequent calculation of the cluster significance probability was done using the Monte Carlo method. The previous steps were repeated once more but calculated for 500 permutations of randomly partitioned data in two subsets. These values were then compared to the observed test statistic. Cluster significance levels were set at a two-tailed $\boldsymbol{\alpha} < 0.025$, and a minimum of two neighbouring channels considered as a cluster. I considered effects falling below $\boldsymbol{\alpha} < 0.025$

as being statistically significant and commented on non-significant (trend-level) effects for clusters that fell between **α** < 0.025 and **α** < 0.1.

4.3.2. Behavioural analyses

The statistical analyses of the probe data and affect ratings were carried out using R (Version 4.2.3; R Core Team, 2018). All subsequently defined linear mixed models (LMMs) were computed using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) R packages.

4.3.2.1. Exploring relations between affect changes and music of different qualities

LMMs were run to check that exposure to music with different qualities had effects on the listeners affect as might be expected. Specifically, I defined a separate model with each affect rating (energy, tension, and valence) as a dependent variable, and with categorical variables Time Point (Pre vs. Post), Familiarity type (Familiar vs. Unfamiliar), and Relaxation Potential type (Relaxing vs. Non-Relaxing) as a fixed effects, as well as participant ID as a random effect.

4.3.2.2. Influence of music qualities and individual differences (musical training, imagery ability, and preference for genres of heard music) on visual imagery incidence rate

Potential differences in the incidence rates (%) of visual imagery occurrence (*yes* vs. *no*) and intentionality (*spontaneous* vs. *deliberate*) as a function of music qualities and individual differences were estimated with LMMs. In a first model, *visual imagery* rates for each participant for each block was indicated as the dependent variable and Familiarity (familiar vs. unfamiliar), and Relaxation Potential (relaxing vs. non-relaxing) as fixed effects with an interaction term specified between them. The second model further included Musical Training, the two Psi-Q dimensions (Vision and Emotion) and mean preference ratings across all used genres (EDM, Classical, Jazz, and Ambient) as fixed effects. Participant was included as a random effect in all models. The model assessing visual imagery intentionality rate was similarly designed, however using occurrences of *spontaneous visual imagery* as the dependent variable.

4.4. RESULTS

Participants reported experiencing visual imagery in response to 72.59% (SD = 22.17) of probes across all listening conditions. Thus, visual imagery was reported more often than *not* experiencing visual imagery (M = 21.15%, SD = 40.85; note that 6.26%, SD = 24.23 pertained to missed probes). Participants also reported experiencing spontaneous visual imagery (M = 70.94%, SD = 45.41) more often than deliberate visual imagery (M = 28.84%, SD = 45.31).

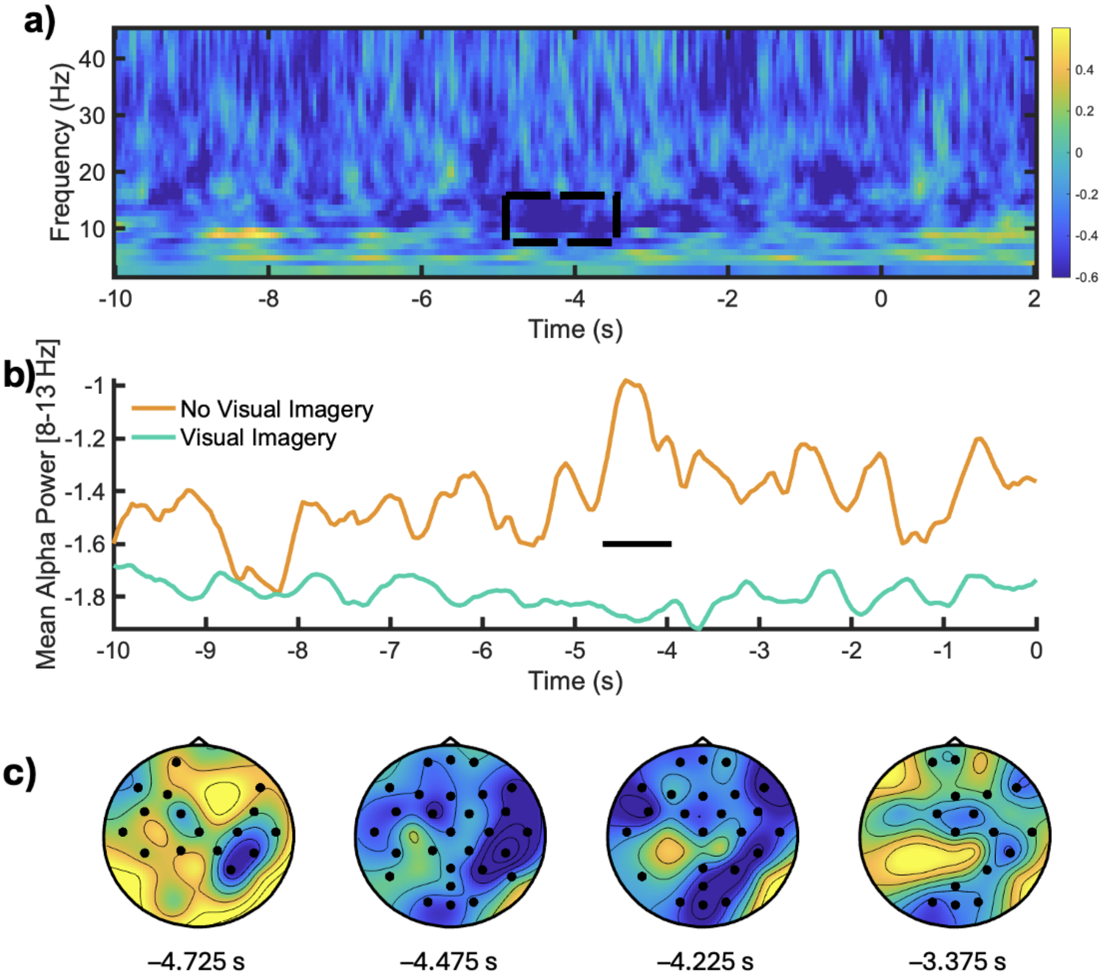
4.4.1. Oscillatory characteristics of visual imagery and its intentionality

4.4.1.1. Alpha band

As can be observed in Figure 4-1, the cluster-based permutation tests revealed a significant suppression of alpha power in the comparison between visual imagery and no visual imagery probes. The analysis revealed this significant difference to occur within a cluster from –4.7 to –3.95 s prior to probe onset, $p = 0.018$. This effect was topographically dispersed, but while it was mainly located around midline fronto-central channels during the initial stages, it moved to right parieto-occipital and midline occipital channels by the end of the window.

Figure 4-1

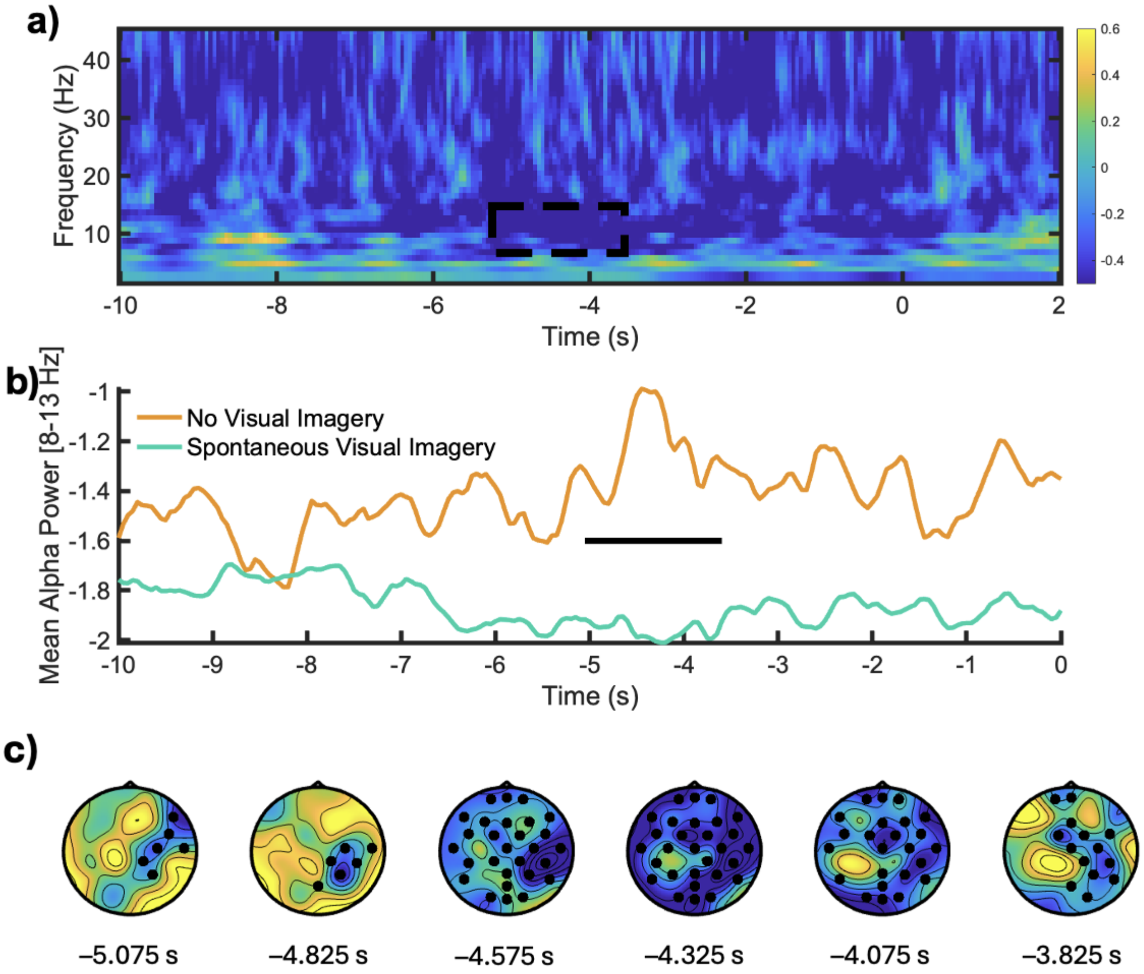
Oscillatory differences between states (visual imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between visual imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Alpha [8-13 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).



The difference between deliberate and no visual imagery states in the alpha band was not significant, though effects tended towards suppressed alpha levels in response to deliberate imagery. However, analyses further revealed alpha suppression between spontaneous and no visual imagery probes in one cluster in the –5.05 to –3.6 s time range, *p* = 0.004 (see Figure 4-2), This cluster spanned bilateral and midline fronto-central sites, but also transiently occupied occipital and right parieto-occipital channels. Trend-level differences were also observed just prior to probe-onset in the –1.05 to 0 s time range, *p* = 0.044, and reached significance with 10 s post-probe baseline (indeed, this baseline showed two significant clusters, *p* = 0.004, in the –5.2 to –3.45 s time range; and *p* = 0.010, in the –1.05 to 0 s time range).

Figure 4-2

Oscillatory differences between states (spontaneous imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between spontaneous imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Alpha [8-13 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).



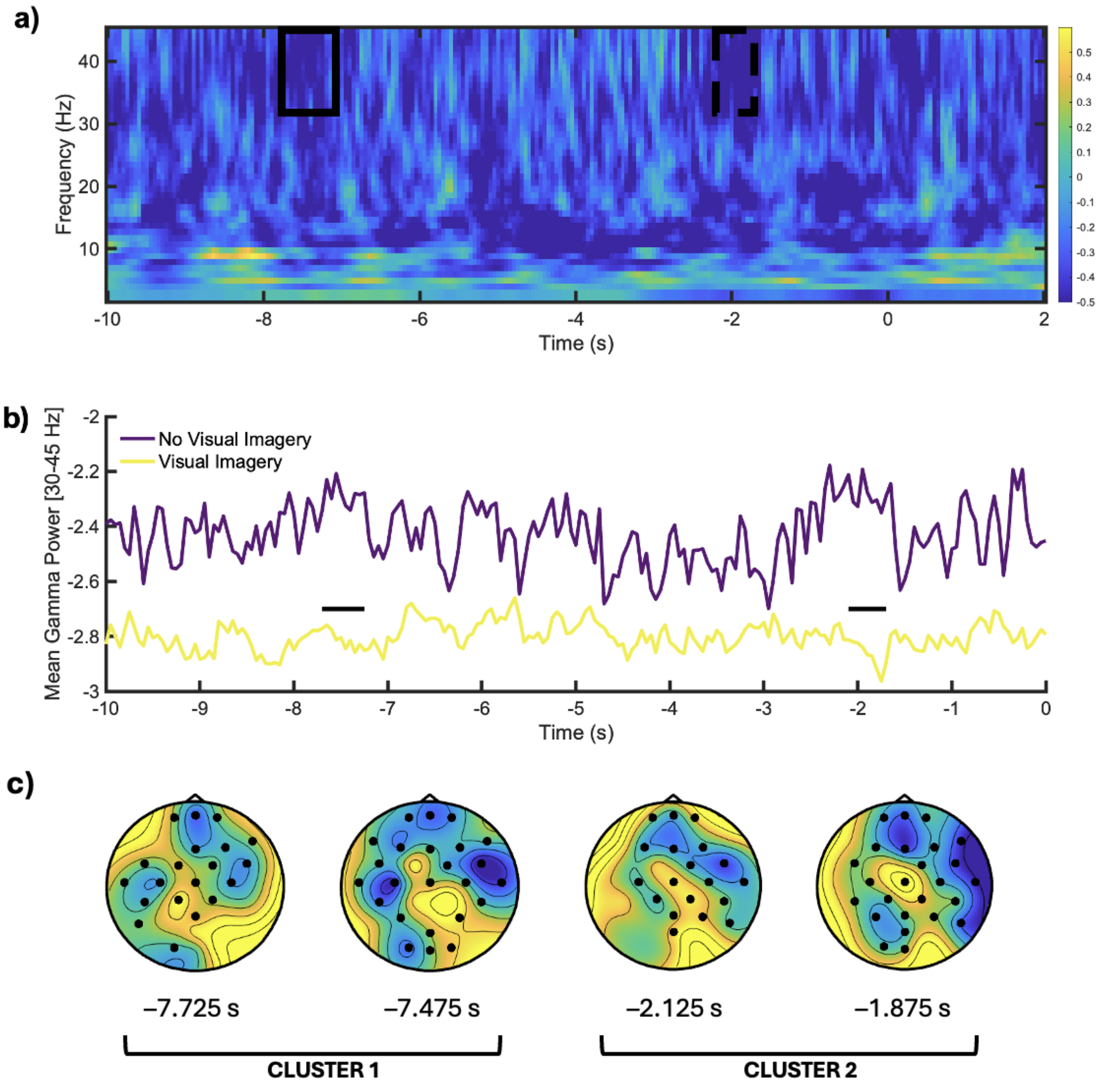
Finally, trend-level statistical differences, p = 0.078 (in the –6.55 to –6.05 s pre-probe time range), that reached significance when using a 10-s baseline (within the 6.7 to 6 s time range, p = 0.018) were seen between spontaneous and deliberate manifestations of visual imagery, whereby there was lesser alpha in right temporal and central channels during spontaneous imagery.

4.4.1.2. Delta, theta, and gamma bands

Analyses revealed no differences between any of the visual imagery states in terms of delta and theta power. However, visual imagery (compared to no visual imagery) was associated with reduced gamma power as visible in two temporally brief clusters (see Figure 4-3). The first cluster (–7.7 to –7.25 s, $p = 0.010$) was mainly observed in midline and bilateral frontal and fronto-central regions (although it transiently occupied midline occipital channels), while the second cluster was closer to probe onset (–2.1 to –1.7 s, $p = 0.010$) and was reflected mostly in frontal and central midline channels, and right and midline parietal channels.

Figure 4-3

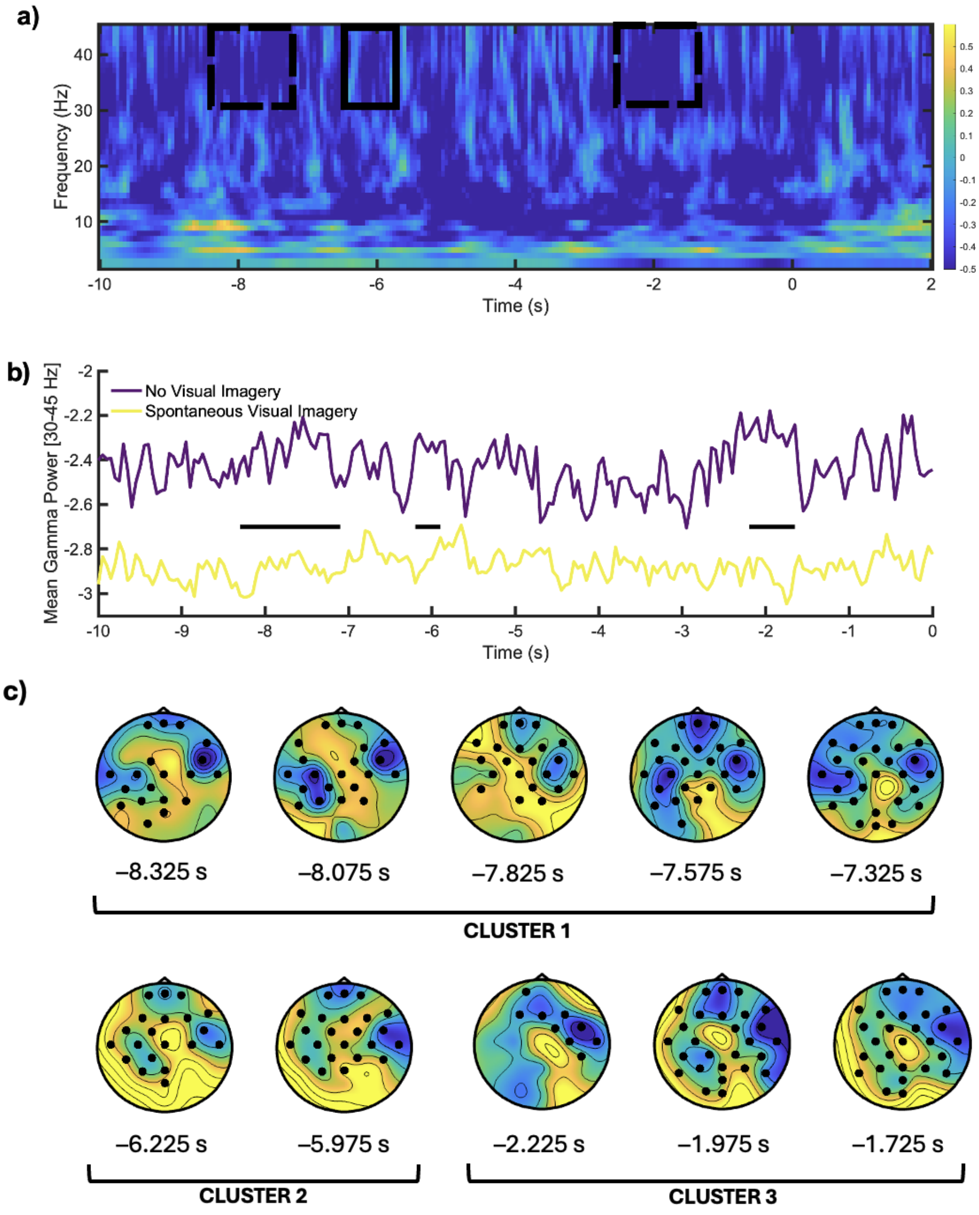
Oscillatory differences between states (visual imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between visual imagery and no visual imagery probes. Significant clusters are indicated by a broken black rectangle. (b) Gamma [30-45 Hz] spectral power averaged over the channels of the clusters. The significant clusters are denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the clusters).



While there was no difference found between deliberate and spontaneous states or between deliberate and no visual imagery states in terms of gamma power, a similar pattern was obtained when comparing spontaneous imagery to no visual imagery states. Specifically, there was less gamma during spontaneous visual imagery found in three clusters (cluster 1: –8.3 to –7.1 s, $p = 0.004$; cluster 2: –6.2 to –5.9 s, $p = 0.020$; cluster 3: –2.2 to –1.65 s, $p = 0.010$), which were most focally localised in midline frontal and fronto-central channels (see Figure 4-4).

Figure 4-4

Oscillatory differences between states (spontaneous imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between spontaneous imagery and no visual imagery probes. Significant clusters are indicated by a broken black rectangle. (b) Gamma [30-45 Hz] spectral power averaged over the channels of the clusters. The significant clusters are denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the clusters).



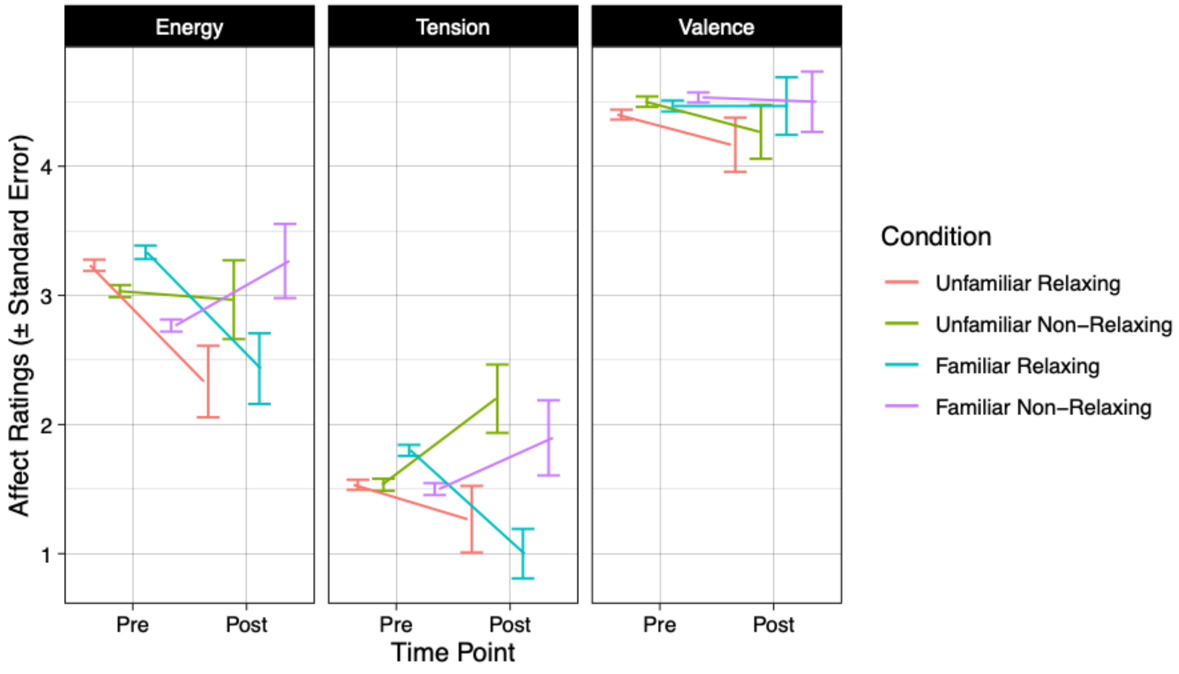
4.4.2. Music stimuli check

LMMs were run to check that exposure to music with different qualities had effects on the listeners as might be expected. See Figure 4-5 for visualisations of relationships separated by each listening block. For Energy, results revealed a significant main effect of time point, $F(1, 203) = 6.39$, $p = 0.012$, and a significant interaction between time point and relaxation potential, $F(1, 203) = 17.08$, $p < 0.001$. Following up this interaction effect revealed that energy ratings in response to the relaxing stimuli were significantly lower post listening (M = 3.28, SD = 1.39) than pre listening (M = 2.38, SD = 1.50), ß = –0.90, SE = 0.18, $t = –5.05$, $p < 0.001$, and that there was no difference in energy ratings between time points for the non-relaxing tracks, ß = 0.22, SE = 0.20, $t = 1.08$, $p = 0.283$. There were no other main effects or interactions, including those pertaining to music differing in familiarity.

For Tension, there was a significant main effect of relaxation potential, $F(1, 202.2) = 8.29$, $p = 0.004$, and a significant interaction between time point and relaxation potential, $F(1, 202.2) = 16.26$, $p < 0.001$. A follow up model showed tension ratings in response to the relaxing tracks to significantly decrease from the pre (M = 1.67, SD = 1.19) to post (M = 1.13, SD = 1.24) time points, ß = –0.53, SE = 0.17, $t = –3.10$, $p = 0.003$. In contrast, tension ratings increased from pre (M = 1.52, SD = 1.33) to post (M = 2.05, SD = 1.50) in response to the non-relaxing tracks, ß = 0.52, SE = 0.19, $t = 2.68$, $p = 0.009$. No other main effects or interactions were found, including those pertaining to music differing in familiarity. Regarding Valence ratings, I did not observe any significant main effects or interaction terms with regards to the effects of exposure to music differing in relaxation potential and familiarity on affect responses between time points.

Figure 4-5

Mean values (and ± standard error of the mean) of affect ratings (Energy, Tension, and Valence) provided before (Pre) and after (Post) taking part in each listening condition (Unfamiliar Relaxing, Unfamiliar Non-Relaxing, Familiar Relaxing, and Familiar Non-Relaxing).



4.4.3. Probe responses and individual differences: characteristics of visual imagery and its intentionality

Linear mixed models predicting the incidence rate of visual imagery showed there to be a significant influence of Familiarity, ß = 0.07. SE = 0.03, *t* = 2.66, *p* = 0.009, and a significant influence of Relaxation Potential, ß = 0.06. SE = 0.03, *t* = 2.12, *p* = 0.036, whereby familiar tracks and relaxing tracks were both associated with higher visual imagery rate. The interaction was not significant, ß = 0.05. SE = 0.04, *t* = 1.45, *p* = 0.150. The analysis looking at spontaneous and deliberate visual imagery alone did not result in any significant main effects or interactions suggesting that neither familiarity nor relaxation potential influenced the rate at which spontaneous imagery is experienced.

In addition, analyses predicting visual imagery and spontaneous visual imagery ratings with regard to their relationships with any of the individual difference measures (musical training, imagery ability, and musical preferences) did not result in any significant main effects or interactions. See Figures C.3 and C.4 in Appendix C for visualisations of these relationships.

4.5. DISCUSSION

The current study used a probe-caught method during EEG recording to investigate the oscillatory characteristics of spontaneous and deliberate music-evoked visual imagery. The main objectives were (1) to replicate previous findings of the neurophysiological patterns of music-evoked visual imagery, and (2) to compare the oscillatory signatures of deliberate and spontaneous visual imagery. Based on past literature (Cooper et al., 2003; Dijkstra et al., 2019; Fachner et al., 2019; Sousa et al., 2017; Xie et al., 2020), I predicted that music-evoked visual imagery would be associated with alpha power suppression in posterior regions of the brain, which I further predicted to be induced in spontaneous imagery to a greater extent than in deliberate imagery. It was further speculated that spontaneous imagery might be associated with enhanced theta and delta power, as for unaware mind wandering (Polychroni et al., 2022), due to the greater likelihood for spontaneous (compared to deliberate imagery) to lie below conscious awareness. Finally, it was predicted that rates of visual imagery would be greater during familiar music (due to a greater potential for evoking autobiographical memories), may vary as a function of relaxation potential (Day et al., 2020; Martarelli et al., 2016), and may be influenced by individual differences.

4.5.1. Signatures of music-evoked visual imagery

In line with several previous studies, including Study 2 (in the previous chapter; see Table 4-1 for a summary of the comparisons between studies) and the one other music-related investigation of visual imagery available (Fachner et al., 2019), I was able to provide evidence for alpha suppression in response to music-evoked visual imagery, and mainly in occipital and parieto-occipital regions responsible for visual processing (Dijkstra, Bosch, et al., 2017; Dijkstra et al., 2019; Xie et al., 2020). Further, spontaneous imagery showed a more reliable and strong alpha suppression than deliberate imagery: this, in line with the prediction that deliberate imagery, due to its top-down nature, may be reported even when less pictorial and vivid, and thus less able to invoke activity in low-level visual regions. However, here it is also very important to point out that there were no differences in degree of posterior alpha suppression when comparing spontaneous and deliberate conditions, suggesting that the differences might not be that great and that deliberate imagery could show more similar patterns to spontaneous in other studies with slightly different conditions.

The results did not reveal any modulation of activity in either theta or delta activity as a function of visual imagery intentionality. Literature on mind wandering had reported elevated theta and delta during unaware types of mind wandering meta-awareness (Polychroni et al., 2022) and is relevant to expect that spontaneous cognition would be more unaware than deliberate. Indeed, given proposals of visual imagery and mind wandering potentially falling under the same umbrella of spontaneous cognition (see Taruffi & Küssner, 2019), and given research showing significant intertwinement between both experiences (Christian et al., 2013; Deil et al., 2023; Taruffi, 2021; Taruffi et al., 2017), I had speculated spontaneous visual imagery would be associated with enhanced theta and delta activity in frontal areas (Polychroni et al., 2022). However, as previously noted, linking intentionality with meta-awareness and visual imagery with mind wandering may be too crude as far as associations go and, as such, expectations of differences may not be well-founded.

In any case, I found visual imagery to be associated with reductions in gamma power and primarily in frontal and central areas. Gamma power in fronto-central areas has been associated with feelings of stress (see Vanhollebeke et al., 2022, for a review) and as such this finding raises the interesting possibility of an association between visual imagery and reduced feelings of stress. It has been suggested that music's ability to reduce stress may be at least partly due to its ability to encourage processes related to autobiographical memory and mental imagery (Panteleeva et al., 2018). However, given that the current study did not manipulate stress, it is clear that further work would be needed to examine such possible relationships.

Table 4-1. Summary of the EEG results presented within the thesis thus far.

Experiment	**Methodology**	**Aims**	**Hypotheses**	**Conditions**	**Outcome Measures**	**Key Findings**
Study 2 (N = 42, Chapter 3)	Retrospective self-report ratings	Examine extent to which static and dynamic music-evoked visual imagery may be associated with differing neural patterns with respect to three regions of interests (frontal, centro-parietal, and parieto-occipital) and three frequency bands of interest (alpha [8–13 Hz], beta [14–30 Hz], and gamma [30–45 Hz]) Explore how imagery-related neural activity differed in key phases of the listening experience (the first and second halves of the piece)	Negative relationship between amount of music-evoked visual imagery and alpha band in parieto-occipital area. Dynamic imagery may involve both beta suppression and enhancement (due to rebound effects) Parieto-occipital gamma enhancement in response to visual imagery generally, but in response to dynamic imagery more specifically	*Imagery Rating Type*: Static Imagery vs. Dynamic Imagery	Self-reports *15-Channel EEG Power*: Alpha (overall, upper and lower), Gamma, Beta	Parieto-occipital alpha **suppression** associated with music-evoked visual imagery Beta **enhancement** to static imagery and **rebound effect** in response to dynamic imagery Gamma **enhancement** to dynamic imagery
Study 3 (N = 30, Chapter 4)	Probe-caught experience sampling	Examine neural correlates of deliberately and spontaneously occurring visual mental imagery in response to music during concurrent EEG recording	Music-evoked visual imagery characterised by alpha power suppression in the posterior regions of the brain Deliberate visual imagery would be less associated with such alpha power suppression (than spontaneous imagery) Spontaneous visual imagery may be associated with increases in theta and delta power	*Listening Condition*: Familiar Relaxing Music Unfamiliar Relaxing Music Familiar Non-Relaxing Music Unfamiliar Non-Relaxing Music	Self-reports *32-Channel EEG Power*: Alpha, Gamma, Theta, Delta	Both visual imagery and spontaneous imagery associated with posterior alpha* and fronto-central gamma **suppression**** Deliberate imagery did not differ from no imagery probes in any frequency band, though also did not show differences from spontaneous imagery in the alpha band No effects found in theta or delta bands

Note. * replication of previous study’s results. ** contrasting with previous study’s results.

4.5.2. Incidence of and factors influencing spontaneous and deliberate imagery

Visual imagery experiences were shown to be prevalent with over 70% of probed moments being associated with imagery (as shown in Study 1 in Chapter 2; Dahl et al., 2022; Küssner & Eerola, 2019; Vuoskoski & Eerola, 2015). This study also provides a window into the rates of types of visual imagery intentionality (Taruffi & Küssner, 2019), showing a higher prevalence of spontaneous visual imagery in comparison to deliberate and no visual imagery rates. This finding coincides with recent evidence that mind wandering most often occurs spontaneously and in the visual domain (Taruffi, 2021).

Past literature that has reported high incidence rates of music-evoked visual imagery has relied on traditional rating methods, which may have encouraged deliberately formed visual imagery. Experience sampling techniques, in contrast, offer the identification of instances in which visual imagery does not occur throughout a listening experience. The results show that the probe-caught methodology, which is often used in mind wandering research (Polychroni et al., 2022; Taruffi et al., 2017), is useful in tracking nuances in the experiences of visual imagery In a study by Taruffi et al. (2017) using a probe-caught paradigm, visual imagery was found to be a prevalent aspect of participants' experiences of music (as, for example, compared to the prevalence of the experience of words – understanding the semantic concept), a result I replicate here. Given the simplicity of the probe-caught method and the effectiveness with which it can measure changes in attentional states (Weinstein, 2018), there is a lot to recommend in exploring the frequency, behavioural, and neural correlates of different types of visual imagery during music listening.

An additional aim of the current study was to relate visual imagery incidence rates with individual differences such as musical training, general abilities to emote and experience visual imagery, and preferences for the music genres used as stimuli. Several previous studies suggest a relationship between visual imagery rates and trait visual imagery to various strengths (e.g., Study 1 in Chapter 2; Hashim et al., 2020; Küssner & Eerola, 2019). This study did not show such an influence of these individual difference measures on (spontaneous and deliberate) visual imagery incidence rates. However, one point to consider is that my approach for examining general visual imagery ability here differs from those used in Study 1 of this thesis and by previous examinations (i.e., the VVIQ; e.g., Küssner & Eerola, 2019), and one possibility is that the scale used here is a less sensitive measure of trait imagery abilities.

Similarly, the lack of a finding of a relationship between imagery incidence and emotion imagination would seem to go against the literature suggesting a close intertwinement of visual imagery and emotion induction (Day & Thompson, 2019; Juslin, 2013; Juslin & Västfjäll, 2008; Vroegh, 2019). However, it remains very possible that these links between emotion and imagery are less strong and extensive than are argued for, and that, indeed, one's general ability to imagine emotions does not necessarily translate into experiencing higher amounts of visual imagery during music listening.

4.5.3. Implications of the research

In addressing research questions surrounding the neural signatures of visual imagery, one important consideration is the demand characteristics that inevitably arise in designs that require participants to explicitly report on their experience. By studying intentionality with a probe-caught method, I was able to investigate for the first time the formation of music-evoked visual imagery in an ecologically valid setting potentially mirroring its natural occurrence in everyday life. This method has presented itself to be important in the understanding of such elusive phenomena.

Here, it is also interesting to note that while visual imagery may be regarded as a form of mind wandering (Taruffi et al., 2017; Taruffi & Küssner, 2019), mind wandering has been associated with alpha enhancements (Polychroni et al., 2022) rather than the alpha suppression reported here, in Study 2 (in the previous chapter), and in several other studies (i.e., Cooper et al., 2003; Fachner et al., 2019; Xie et al., 2020). Previous literature has proposed alpha enhancement during mind wandering to reflect diminished perceptual processing and a detachment from the external world (Polychroni et al., 2022), ideas in line with the concept of cognitive 'idling' and reduced attentional processing (Cooper et al., 2003; Foxe & Snyder, 2011; Palva & Palva, 2007) and different from the emphasis on alpha as tracking more or less visual processing when found in the occipital lobe. Nevertheless, future work is necessary to better understand both the extent of the overlap and distinctions between visual imagery and mind wandering, and the extent to which observed alpha power patterns are reflective of the different processes.

Finally, previous research has proposed the potential for visual imagery to facilitate wellbeing and act as an important tool in therapeutic settings. For instance, deliberate imagery was explored in a study by Herff et al. (2022) where participants, while presented with music, imagined the continuation of a figure's journey towards a landmark. They found that their paradigm affected the emotional sentiments of the deliberate visual imagery generated from a combination of music and eye closure. Their study poses interesting clinical implications that the control and moulding of imagery could have for emotional and self-regulation. Interestingly, the finding of reduced frontal gamma power during visual imagery raises the possibility that visual imagery may indeed be associated with reductions in a brain signature associated with stress.

4.5.4. Conclusion

The findings of Study 3 corroborate past literature linking posterior alpha suppression with music-evoked visual imagery, while expanding it to show how neuro-oscillatory correlates of visual imagery differ as a function of intentionality. The observed effects hint at the relevance for more careful measurements into forms of visual mental imagery that are not simply a result of explicit questioning, and further expand upon the lack of neuroscientific investigations into music-evoked visual imagery. However, it was unable to show if and how visual imagery may play a role in modulating affective states in the listener. In the following chapter, I use EEG, once more, alongside physiological and self-report methods, to explore visual imagery's purported role as a mechanism for stress reduction during music listening.

CHAPTER 5. IS MUSIC-EVOKED VISUAL IMAGERY A VIABLE STRATEGY FOR STRESS REDUCTION?

Music's role as a tool for mood regulation is widely documented in the literature. Alongside this body of work, visual imagery has been suggested as a key mechanism by which music is able to influence a listener's affective state. However, evidence that music-evoked visual imagery – whether deliberately or spontaneously formed – can reduce negative affect, over and beyond said music's acoustic features remains to be examined. To this end, Study 4 used subjective ratings, recordings of electroencephalography (EEG) and skin conductance response (SCR) to investigate whether music-evoked visual imagery can be considered an effective tool for stress reduction. Using a probe-caught paradigm, thirty participants experienced a multi-component stress induction task before listening to relaxing music, non-relaxing music, and a radio show podcast (active control listening track) across three blocks. State anxiety measurements were taken before and after stress induction, as well as after listening to the different tracks. Findings suggested that music-evoked visual imagery is associated with enhanced stress reduction over and beyond the acoustic features of the music. Further, findings that visual imagery, particularly spontaneous imagery, was associated with posterior alpha suppression and fronto-central gamma were replicated. The study also revealed delta suppression during visual imagery, as well as intentionality-specific theta modulation patterns: suppression in the context of deliberate imagery and enhancement in response to spontaneous imagery. The findings provide evidence that visual imagery during music listening can have therapeutic benefits in anxiety and stress-related states.

5.1. INTRODUCTION

In contemporary society, stress has evolved into a pervasive societal concern, extending far beyond individual experiences to infiltrate various aspects of collective existence. Its influence permeates economic structures and social dynamics, impacting workforce productivity and community cohesion alike. As stress continues to exert its far-reaching influence on economies and societies, a paradigm shift towards non-pharmacological interventions is imperative, with various solutions increasingly being explored within psychological research.

Across various non-pharmacological interventions aimed at mitigating stress, music emerges as a compelling avenue for exploration, offering unique therapeutic potential that has been repeatedly demonstrated within research. Numerous studies have elucidated the impact of music on human emotions, cognition, and physiological responses, suggesting its efficacy as an effective tool in stress reduction. Music's capacity to modulate mood states, evoke memories, and induce relaxation has been extensively documented, with empirical evidence supporting its ability to alleviate symptoms of anxiety and depression (Chan et al., 2011; Davis & Thaut, 1989). Moreover, the therapeutic effects of music extend beyond auditory stimulation, and encompasses the potential beneficial effects of visual imagery evoked by the musical stimuli. The exploration of music-evoked visual imagery as an underlying mechanism for stress reduction represents a promising avenue for research, as it can encompass soothing and vivid mental images that may aid in redirecting attention (Herff et al., 2023; Küssner & Eerola, 2019). To this end, the current study explores the potential stress alleviating effects of visual imagery evoked during extended listening.

5.1.1. The psychology and neuroscience of stress and anxiety

Stress and anxiety are intertwined in their behavioural and neural underpinnings with regards to their physiological responses, cognitive processes, and emotional states (Bremner et al., 1996; Daviu et al., 2019; Woody et al., 2018). Stress, defined as the body's response to perceived threats or challenges, triggers a series of physiological changes governed by the sympathetic nervous system and the hypothalamic-pituitary-adrenal (HPA) axis (Cacha et al., 2020). Upon encountering a stressor, the brain initiates the release of stress hormones, including cortisol and adrenaline, which mobilise the body's resources to cope with the perceived threat. This physiological response, commonly known as the "fight-or-flight" response, primes the individual for action, heightening alertness, increasing heart rate, and redirecting energy towards essential functions.

However, chronic or excessive activation of the stress response system can lead to several adverse health outcomes, including cardiovascular disease, immune dysfunction, and mental health disorders such as anxiety and depression (Kemp & Felmingham, 2008). Anxiety, characterised by persistent feelings of apprehension, worry, and tension, represents a maladaptive response to perceived threats or uncertainties, often accompanied by physiological arousal. Neuroimaging studies have elucidated the neural circuitry underlying anxiety disorders, implicating regions such as the amygdala, prefrontal cortex, and hippocampus in the regulation of emotional responses, threat detection, and fear conditioning (Cacha et al., 2020; Daviu et al., 2019; Hölzel et al., 2010; Nolte et al., 2011).

Increases in stress and anxiety have been shown to be indexed by skin conductance (Lazarus et al., 1963), cortisol (Woody et al., 2018), and, to some extent, oscillatory (alpha, beta and gamma frequency bands) neural activity in the brain (Ehrhardt et al., 2021; Minguillon et al., 2016). However, while such physiological indices are important in offering objective measures of these negative states, they are best complemented with measures of subjective experiences that can be captured using validated psychometric tools (questionnaires). The State-Trait Anxiety Inventory (Spielberger et al., 1983) is a validated tool with which to examine current feelings of anxiety as well as general inclinations toward anxiety. Given the close interlink that stress and anxiety are purported to have, collecting state measurements of anxiety is likely to be a useful indicator of the adverse outcomes of an acute stressor, as well as the subsequent reduction in stress caused by a stress alleviating intervention.

The chronic wide ranging negative side-effects that stress-related anxiety can have on an individual's emotional and physical states reflects the idea that there is a need for more non-pharmacological interventions that are inexpensive and accessible. There are several interventions that exist that target the emotional and physiological outcomes of stress-related anxiety, such as mindfulness (with varying effectiveness; S. R. Bishop, 2002; Grossman et al., 2004) and breathing exercises (Decker et al., 2019; Jerath et al., 2015) for self-regulation. However, there is promising evidence on the positive effects of music listening that suggests it might have the capacity to decrease psychobiological stress response.

5.1.2. Music and stress reduction

Music has long been explored as a technique for reducing feelings of stress and anxiety (Baltazar et al., 2019; Davis & Thaut, 1989). Indeed, a large body of research has shown that music holds a profound ability to modulate emotional state and mood (Rickard, 2004). Music's use as a strategy for reducing stress (Yehuda, 2011) seems to be driven in large part by listeners' recognition of the inherently relaxing qualities of certain features of music (Baltazar et al., 2019). Specifically, when using music to destress or reduce anxious feelings, listeners are usually drawn towards features such as slow tempo, limited dynamic variation, and rhythmic simplicity (Tan et al., 2012).

Specifically, music of varying styles has been shown to affect subjective and physiological measures of stress. Across two studies, Groarke et al. (2020) showed that self-selected music was effective in reducing self-reported anxiety and blood pressure after an acute stressor (when compared to silence), as well as increasing state mindfulness. Music also has additional benefits that can target the wider psychobiological stress system, including cortisol levels. Thoma et al. (2013) found that relaxing music (when compared to the sound of rippling water and to unstimulated rest) led to faster reductions in cortisol and salivary alpha-amylase levels to baseline when listened to prior to undertaking the stress induction task. Five minutes of music listened to prior to stress induction has also been found to decrease heart rate more effectively than when listened to after stress induction (G. Chen et al., 2019), suggesting that music constitutes a powerful preventative measure than a remedial one (K.-C. Lee et al., 2012). Finally, in a recent meta-analysis by Harney et al. (2023), it was concluded that music listening interventions, whether self-selected or experimenter-selected, can reduce state anxiety with high effect sizes.

However, it has been argued that music may offer relaxation in ways that are not strictly tied to intrinsic features of the music. Notably, a meta-analysis by Panteleeva et al. (2018) proposes that in addition to physiological effects that may occur due to music's structure and acoustic features, more complex cognitive mechanisms such as spontaneous autobiographical memories and mental imagery, may also play a critical role in music-related stress and anxiety reduction. Given the therapeutic implications of finding music-evoked imagery to be associated with stress reduction, the current study sought to examine this possibility, using both subjective and (electro)physiological measures of imagery and affect.

5.1.3. Music-evoked visual imagery as a stress reduction strategy

Recent behavioural research into the content and function of visual imagery suggests that music can induce visual imagery with the ability to soothe (Küssner & Eerola, 2019). However, evidence for the role of visual imagery in stress induction is still limited. Using the BRECVEMA mechanisms (Juslin, 2013; Juslin & Västfjäll, 2008), Baltazar and Västfjäll

(2020) estimated the cumulative influence of each of the eight mechanisms (*brain stem reflexes*, *rhythmic entrainment*, *evaluative conditioning*, *emotional contagion*, *visual imagery*, *episodic memory*, *musical expectancy*, *aesthetic judgement*) in terms of their contributions towards relaxation. Specifically, an online participant pool was asked to provide tracks that they had previously used to calm themselves in stressful situations as well as tracks that would not work for them in this situation before then rating a list of mechanisms with regard to their contribution to the track's relaxation potential. The authors extracted the tracks' audio features, including danceability, loudness, and tempo. Interestingly, the authors found that key mechanisms for inducing relaxation included the track's aesthetic appeal, genre, and familiarity, and that key influential audio features included (in order of descending influence) energy, loudness, valence, acousticness, and speechiness. They observed that visual imagery, in contrast, ranked much lower as a mechanism by which music is relaxing.

Here, it is important, however, to note that Baltazar and Västfjäll's study focused on the particular *strategies* that individuals actively employ when their aim is to relax. Thus, one possibility is that even though visual imagery may not be a foremost mechanism utilised across the general population and across a range of everyday contexts, visual imagery may still serve a relaxation function 1) particularly in people with a high capacity for visual imagery; and 2) particularly in contexts where listeners can carry out active listening and imagery without having to engage in other tasks (e.g., driving, studying).

5.1.4. Neural correlates of music-evoked visual imagery

In addition to this, Studies 2 and 3 (Chapters 3 and 4, respectively) have helped to establish, amongst music- and non-music-related literature (e.g., Dijkstra et al., 2019; Fachner et al., 2019; Sousa et al., 2017; Xie et al., 2020), posterior alpha suppression as a reliable index of music-evoked visual imagery content that may be more present during spontaneous, than deliberate imagery: this, possibly due to deliberate imagery containing less pictorial and vivid content, and thus invoking activity in lower-level visual regions (Dijkstra, Zeidman, et al., 2017; Gilbert & Li, 2013; Stokes et al., 2009). My previous study also implicated frontal gamma band suppression in visual imagery. Gamma has been found to be suppressed in instances of internally-directed attention (Lehmann et al., 2001; Villena-González et al., 2018), but enhanced in moments of high task complexity and load (De Lange et al., 2008; Filgueiras et al., 2018; Han et al., 2022; Sepúlveda et al., 2014) and stress (Ehrhardt et al., 2021; Minguillon et al., 2016).

It nevertheless remains to be seen whether such effects can be replicated. Furthermore, the previous study did not find effects with regard to the theta and delta frequency bands, begging the question of whether music-evoked visual imagery and meta-awareness of mind wandering may not always be comparable or whether such neuro-oscillatory patterns may play a role in imagery. Indeed, suggesting that they might, a study by Benedek et al. (2011) found that increased theta power in EEG was associated with enhanced visual imagery ability. Further, a study by Alluri et al. (2013) found increased theta synchronisation in EEG when participants imagined scenes related to music. Also, both theta and delta have been implicated in memory processes, including encoding and retrieval (Klimesch, 1999; Rasch et al., 2007). Thus, while arguments associating these bands to music-evoked visual imagery are still speculative, the current study further explores whether theta and delta could show to differentiate between imagery states or not.

5.1.5. The current study

Using the probe-caught methodology again (Polychroni et al., 2022; Taruffi et al., 2017) during concurrent EEG (Compton et al., 2019; Liu et al., 2021), the current Study 4 examined the potential role of visual imagery in explaining music's ability to reduce signatures of stress following a stress-inducing task. An additional aim was to explore whether spontaneous and deliberate visual imagery would differ in their propensity for stress reduction. To test the efficacy of music-evoked visual imagery in aiding in stress reduction, stress was induced using an adapted version of the 4-component Mannheim Multicomponent Stress Test (MMST; Reinhardt et al., 2012), comprising auditory, emotional, cognitive, and motivational stressors.

Subjective experience of stress-related anxiety was measured using a reliable tool for measuring self-reported anxiety response (the State-Trait Anxiety Inventory; STAI-6; Marteau & Bekker, 1992) while objective experience was measured using skin conductance response (SCR). Finally, to obtain evidence of a role of music-evoked visual imagery that goes beyond the acoustic features of the music, three conditions were used: namely a musical track with high relaxation potential due to its low energy, low danceability, and low loudness features; a track with low relaxation potential due to its high energy, high danceability, and high loudness features and an active auditory (non-musical) control track. A further exploratory aim was to examine the extent to which general imagery abilities related to visual imagery and emotion, as well as trait features of anxiety and stress reactivity, could play a role in influencing the magnitude of drops in subjective and physiological stress-related anxiety.

The hypotheses of the current work were as follows: (1) music-evoked visual imagery would be related to posterior alpha power suppression, with spontaneous imagery showing this activity pattern to a greater extent than deliberate imagery; and (2) while music with features promoting relaxation will be associated with drops in objective and subjective signatures of stress, the incidence of visual imagery will also be shown to be predict the size of these drops in signatures of stress.

5.2. METHODS

5.2.1. Participants

Data was collected from 30 healthy adults, aged 18-59 (22 female, 8 male; M = 25.0, SD = 8.6) who were recruited using poster advertisements and word-of-mouth. An online survey was used to assess eligibility for the research. Specifically, the survey required participants to report (on a 1-7 Likert scale, from 1 = ‘Not at all’ to 7 = ‘Experience very strongly’) on the extent to which they experienced the following phenomena in response to music: chills or goosebumps, visual imagery, induced emotions, and personal memories (a selection of experiences were asked in order to help mask the intentions of the study). The primary inclusion criterion was that individuals should report experiencing at least mild levels (providing a rating of 2 or upwards) of visual imagery in response to music. Those who reported never experiencing music-evoked visual imagery (i.e., provided a rating of 1) were considered ineligible and were thanked for their time. As part of this screening process, participants were also advised that those with a history of a diagnosis of post-traumatic stress disorder, depression and/or anxiety should consider their participation carefully, due to the emotionally charged stimuli that would be included in the experiment. These stimuli referred especially to the emotional stressors used in the stress induction task (see *Stress induction* section in *5.2.3 Materials and stimuli* below).

5.2.2. Ethics statement

The research was approved by the Ethics Committee of the Department of Psychology at Goldsmiths, University of London. All participants provided online and in-person written informed consent before taking part. Eligible participants were provided with monetary compensation of £10 per hour for their time.

5.2.3. Materials and stimuli

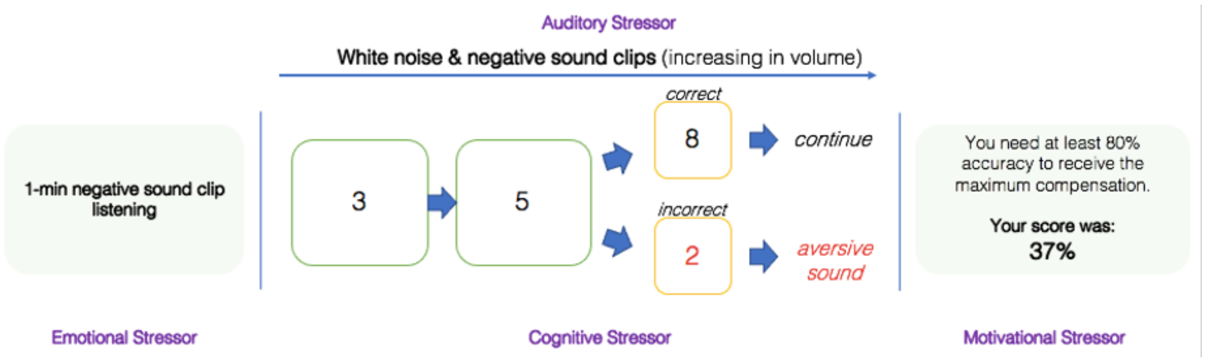
Listening conditions. The listening task comprised three conditions that I labelled structurally relaxing music, structurally non-relaxing music, and an active control (a radio show). The structurally relaxing music was an Ambient track, intended to elicit low arousal levels. The structurally non-relaxing track was an upbeat Techno track matched for genre and which was chosen from the pilot survey presented in Study 3 (in the previous chapter) as being rated highly in energy, valence, danceability, loudness. An active control Podcast condition was selected over a silence condition due to the impracticalities of asking participants to experience 15 mins of silence and thus most likely high levels of stress and boredom. This control condition was a 15-min segment from the radio show 'A New Take on Darwin's "Origin of Species"', which was used in a previous study that compared the effects of self-selected and researcher-selected music on induced negative affect (Groarke et al., 2020).

Stress induction. Stress was induced using a 5-min modified version of the Mannheim Multicomponent Stress Test (MMST; Reinhardt et al., 2012), a task combining four types of stressors: *cognitive* (mental arithmetic), *auditory* (white noise), *emotional* (affective pictures), and *motivational* (loss of money). See Figure 5-1 for an outline of this task. In my adapted version, the affective pictures for the *emotional* stressor were substituted with affective sound clips to avoid any confounding effects that the affective pictures could later have on visual imagery experience during the listening task. Affective sound clips were obtained from the International Affective Digitised Sounds database (IADS; Bradley & Lang, 2007). Negative sound clips comprised excerpts that were rated lowest in pleasure and highest in arousal, whereas positive clips comprised those that were rated highest in both pleasure and arousal. Sound clips were played consecutively for the first minute. After each block of five sound clips with negative valence, a clip with positive valence was presented to avoid habituation effects. Each sound clip was about 3 s long. Two of the sound clips were demonstrated twice. In order to maintain attention, participants were asked to verbally identify the second iteration as soon as it occurred.

All remaining MMST components were implemented as in its original version. The *cognitive* stressor was a modified computer version of the Paced Auditory Serial Addition Task (PASAT-C; Lejuez et al., 2003). For the remaining 4 mins, numbers from 1 to 9 were sequentially flashed on a screen and participants were asked to sum the most recent number with the previous number. Using a series of boxed numbers on the screen ranging from 1 to 18, they were instructed to click on the correct answer. After providing each sum, the participant had to ignore the sum they had just provided and add the following number to the most recently presented number. Sound clips from the *emotional* stressor continued to play in the background, with a 5-s inter-stimulus interval. The white noise for the *auditory* stressor played throughout the duration of the 5-min induction task. The intensity of the white noise was increased from 78 to 93 dB to avoid habituation. For the *motivational* stressor, participants were told that they were required to reach an accuracy score of at least 80% or higher (presented to them after stress induction) on the *cognitive* stressor to be eligible for full compensation. The difficulty of the *cognitive* stressor was designed to reach close to 80% with high effort responses (to not demotivate the participant), though to never genuinely reach it. Participants were debriefed at the end of the study that their compensation was in reality not affected by the task.

Figure 5-1

The stress induction task (modified MMST), which comprised four types of stressors: emotional (negative sound clips), auditory (white noise and negative sound clips), cognitive (arithmetic task), and motivational (loss of money).



Probe-caught measures. Visual imagery experience was assessed by presenting thought probes at pseudorandom intervals throughout the listening trials as in the previous study. Probes were presented every 40 s with a uniform ± 10-s jitter (i.e., all possible durations between 30 and 50 s across every listening condition). Phrasing and implementation of the thought probes followed previous MW literature (Polychroni et al., 2022; Taruffi et al., 2017; Weinstein, 2018). The first probe prompted the participant to indicate if they were experiencing any visual imagery: '*Just before the probe, were there any images in your mind's eye?*', presented with the dichotomous response boxes: '*Yes*' and '*No*'. If the participants responded with '*Yes*', they were asked a follow-up question: '*Were your thoughts deliberate and actively generated or spontaneous and free-flowing?*', presented with the dichotomous response boxes:

'*Deliberate'* and '*Spontaneous'*. Participants were given up to 4 s to answer each probe. These parameters were set to allow presentation of as many probes as possible (a total of up to 25 probes per condition and up to 100 probes over the four conditions), while allowing sufficient time in between probe presentations for the participant to return to the listening task. This was also to discourage participants from overthinking their answers and to avoid interrupting the listening experience for too long.

Behavioural ratings. Affective state was obtained using the shortened version of the State Trait Anxiety Inventory (Spielberger et al., 1983), the STAI-6 (Marteau & Bekker, 1992), consisting of the following terms from the original inventory: '*I feel calm*', '*I am tense*', '*I feel upset*', '*I am relaxed*', '*I feel content*', and '*I am worried*'. This was presented before and after the stress task and was also probed in three instances throughout the listening experience (i.e., after every 8 visual imagery probes, including one immediately after the end of listening). Additional questions inquiring into the content of visual imagery experienced, as well as general impressions regarding the tracks (see Table C.3 in Appendix C), during the listening were also presented at the end of each block, including an open-text question asking for descriptions of visual imagery content. However, these will be analysed elsewhere.

Sham trials. A sham task was included at the end of each block to deflect the participant from the true aims of the experiment. This comprised three trials, where the participant would select a shape and colour combination that they preferred most out of the options shown. This is a simplified version of a task developed on the idea proposed by Wassily Kandinsky that individuals have implicit associations between colour and form, a theory that has yielded inconsistent results (Braun & Doerschner, 2019; N. Chen et al., 2015; Dreksler & Spence, 2019; Makin & Wuerger, 2013).

Individual differences. The Perceived Stress Reactivity Scale (PSRS; Schlotz et al., 2011), a 23-item questionnaire, was used to assess individual differences in levels of perceived stress reactivity. In this questionnaire, individuals are presented with stressful scenarios and asked to indicate using three potential options (varying in reactivity) what their typical reaction would be in response.

Trait anxiety was measured using the Form Y-2 of the State Trait Anxiety Inventory (STAI-T; Spielberger et al., 1983), a 20-item self-evaluation questionnaire where individuals are asked to describe how they generally tend to feel according to a series of statements. Each statement was rated on a 4-point Likert scale from 1 = 'Almost Never' to 4 = 'Almost Always'.

The 7-item Musical Training subscale of the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) was presented to gauge participants' levels of prior musical training (e.g., 'I have never been complimented for my talents as a musical performer', 'I am able to judge whether someone is a good singer or not'), using a 7-point Likert scale from 1 = 'Completely Disagree' to 7 = 'Completely Agree'. Finally, I presented the 5-item Vision and 5-item Emotion subscales of the Plymouth Sensory Imagery Questionnaire (Psi-Q; Andrade et al., 2014), which assesses one's general ability to experience mental imagery across multiple modalities (such as visual, tactile, emotion, olfactory, etc.; e.g., 'Imagine the appearance of … a sunset', 'Imagine feeling … excited'). These items were rated on an 11-point Likert scale from 0 = 'no image at all' to 10 = 'imagery as clear and vivid as real life'.

5.2.4. Procedure

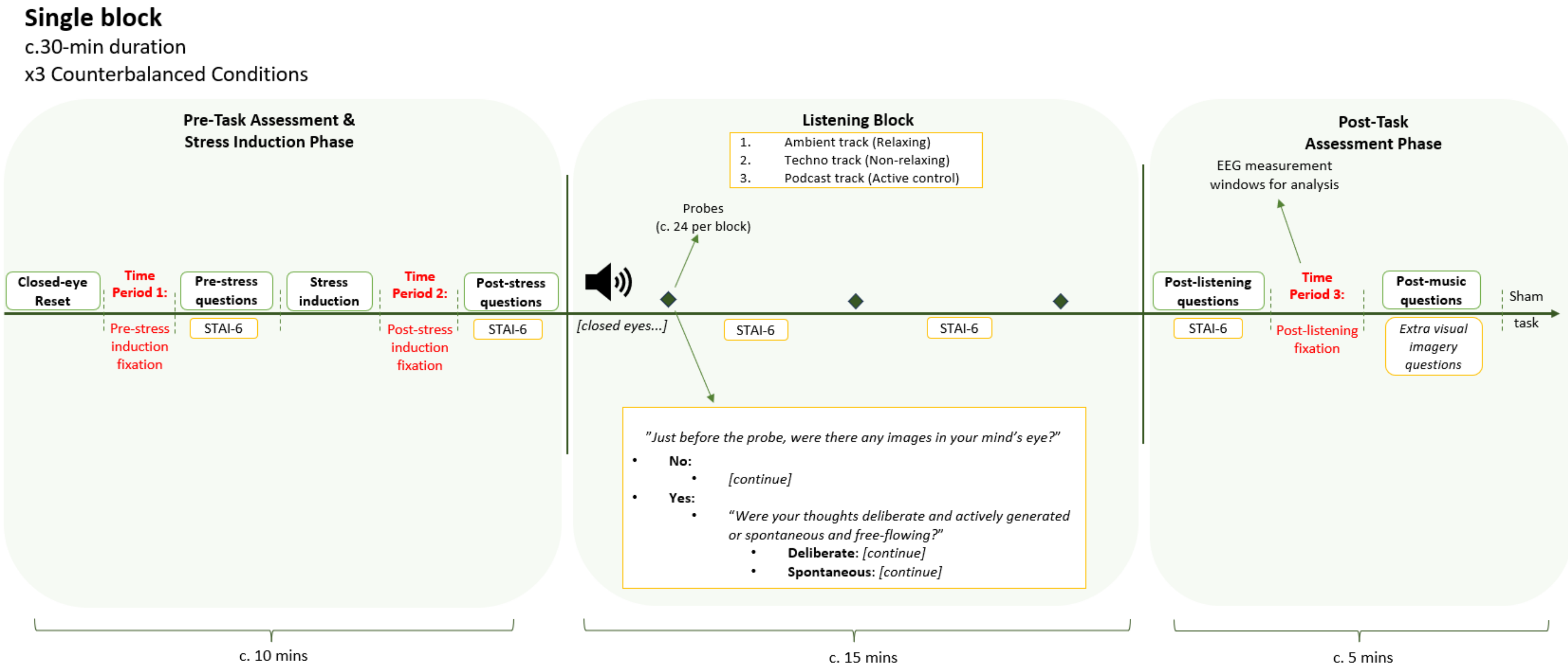
The experiment was programmed and presented using OpenSesame (Version 3.3.8, Python 3.7.9; Mathôt et al., 2012). Participants were fitted with the EEG and the SCR. They sat in a dimly lit room and were presented with verbal as well as written instructions on the computer screen for the experiment. The stimuli were played through headphones and participants were advised to adjust the volume to a comfortable level and maintain that volume for the duration of the experiment. Participants completed three blocks, and the duration of the experiment was approx. 90 mins. Prior to the start of the experiment, deliberate visual imagery was defined as visual thoughts that are purposefully thought out with purpose, whereas spontaneous visual imagery was defined as appearing without any particular intention or control. Participants first underwent an easier and shorter practice version of the stress task, before being shown a demo of the probe questions.

Participants first sat in silence with eyes closed for 60 seconds (intended to neutralise mental state at the start of each block), followed by the STAI-6, then a 60-second fixation that allowed sufficient skin conductance data to be collected without movement artefacts. Next, participants completed the 5-min amended MMST stress induction task (see Figure 5-1) before once more completing the STAI-6 and fixating for 60 seconds again. For the 15-min listening task that followed, participants had been informed that they should listen to the track, immersing themselves with eyes closed, and that they would hear a brief clicking sound whenever a probe question had appeared on the screen. At that point, participants were required to provide their answer(s) to the probe(s) and then continue to listen with their eyes closed.

At the end of each block, participants completed the STAI-6, fixated for 60 seconds, and then answered the fifteen additional questions about the nature of their visual imagery experience. To finish, participants completed the sham task. See Figure 5-2 for a summary of the procedure of a single block.

Figure 5-2

Experimental flow of a single block.



5.2.5. Skin conductance response (SCR) and EEG recording

Physiological measurements of SCR and heart rate (HR) were measured using Shimmer3, a wearable sensor device whereby small electrodes were placed and strapped onto the fleshy bottom parts of the index and middle fingers to measure SCR, and on the tip of the index finger to measure HR, although only SCR is analysed and reported here. The electrodes were connected to a small measurement device that would record the SCR remotely, which was placed in the palm of the participants' non-dominant hands for the duration of the experiment. Participants were encouraged to not overly squeeze or move their hand during recording. SCR was recorded at a sampling rate of 512 Hz.

EEG was recorded using the mobile Waveguard 32-channel cap from ANT Neuro, The Netherlands. The 32 channels used were: Fp1, Fpz, Fp2, AFz, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, M1, CP5, CP1, CP2, CP6, M2, P7, P3, Pz, P4, P8, POz, O1, Oz, O2. The cap was placed in accordance with ANT Neuro's 5% electrode placement system, a derivative of the 10-20 international system. Signals were amplified using the eego EEG recording amplifier, with the reference set to electrode CPz. Impedance levels were always kept below 20 kΩ. The EEG recording was set to a sampling rate of 500 Hz for all participants.

5.2.6. Skin conductance responses (SCR) analysis

The SCR data was epoched into the 60-second time points of interest: Pre-Stress, Post-Stress, and Post-Listening. Each participant's epoched data was then pre-processed by applying a low-pass Buttersworth filter between the frequencies 1 and 50 Hz to remove line noise. A moving average was then applied to the filtered data, before being centred and scaled. The averaged time point data was then subdivided into their corresponding listening block for later statistical analyses.

5.2.7. EEG analysis: preprocessing and epoching

The EEG data was imported into MATLAB using EEGLab (Delorme & Makeig, 2004), and pre-processed and further analysed using functions from the FieldTrip toolbox (Oostenveld et al., 2011). Data segmentation, pre-processing and analysis choices were broadly in line with previous studies (Compton et al., 2019; Liu et al., 2021; Polychroni et al., 2022). The continuous data was first segmented into 34.7-s epochs: –12 s before to 22.7 seconds after probe onset. The time window of interest was from –10 to 0 s relative to probe onset. However, the longer epoch was taken to allow a 700-ms baseline window within the interstimulus interval after probe onset and the start of the next probe trial, as well as to avoid edge artefacts.

The segmented data was then resampled to a rate of 200 Hz with a low-pass band filter set at a frequency of 50 Hz and a notch filter set at a frequency of 48-52 Hz. Bad or faulty electrodes were first visually searched for by observing the raw data, but as none were found, no electrodes were removed or interpolated in this step. Artefacts were then visually assessed by scrutinising the time courses of each trial's raw data and replacing bad segments with NaN's (average seconds removed = 1.45, max = 4.63, min = 0.005). Next, independent component analysis was used to identify and remove artefacts caused by eye movements, eye blinks, heartbeats, and further channel noise (average number of components removed = 4.7). Finally, trial ✕ channel spectra were visually assessed to reject whole trials that still exhibited high levels of variance even after prior artefact removal efforts, however no trials were removed in this step. All the data was then re-referenced using the average of all the channels.

Finally, the epoched data was subdivided into four conditions corresponding to all possible visual imagery states reported by the participants: *no visual imagery*, *visual imagery*, *deliberate* visual imagery, and *spontaneous* visual imagery, intended for later cluster analyses.

5.2.8. EEG analysis: time-frequency decomposition and statistical analysis

A time-frequency decomposition of the data was computed using a Hanning taper at 50 ms steps along the 34.7-sec epoch, including the 700 ms baseline window for frequencies from 1 to 50 Hz. This was carried out using a frequency-dependent window length, with 7 cycles per time window. Spectral power for every trial was averaged then normalised by dividing it by the baseline level (i.e., the gain model; Grandchamp & Delorme, 2011).

The 700-ms baseline was selected from 15 to 15.7 seconds post probe onset, as it was late enough in the inter-stimulus interval to be sure that participants had closed their eyes again after responding to a probe (participants had up to 8 seconds to respond to the two questions) but early enough to avoid that substantial visual imagery had been formed again. In any case, comparing this choice of baseline period against one 10 seconds and another 20 seconds after probe onset showed that the nature of effects did not differ substantially across the three types. Any discrepancies will be highlighted throughout the results.

Differences in the spectral power between the different visual imagery states (visual imagery vs. no visual imagery, spontaneous vs. deliberate, spontaneous vs. no visual imagery, and deliberate vs. no visual imagery) were analysed using non-parametric cluster-based paired permutation tests (Maris & Oostenveld, 2007), run for each of the oscillatory frequency bands of interest (delta [2-3 Hz], theta [4-7 Hz], alpha [8-13 Hz], and gamma [30-45 Hz]) for completion. These were carried out using the Fieldtrip toolbox (Oostenveld et al., 2011) in MATLAB.

These analyses were first carried out by calculating the observed test statistic by means of paired samples t-tests between the specified contrasts at each sample of the 3-dimensional data (channel vs. frequency vs. time). The samples for which the t-statistic fell beneath the $\boldsymbol{\alpha} < 0.05$ threshold were selected and clustered in connected sets on the basis of temporal vs. spatial vs. spectral adjacency. The t-statistics within every cluster were then summed to calculate the cluster-level statistics. The maximum of these statistics served as the test statistics with which differences between visual imagery and manifestation states were determined.

Subsequent calculation of the cluster significance probability was done using the Monte Carlo method. The previous steps were repeated once more but calculated for 500 permutations of randomly partitioned data in two subsets. These values are then compared to the observed test statistic. Cluster significance levels were set at a two-tailed $\boldsymbol{\alpha} < 0.025$, and a minimum of two neighbouring channels considered as a cluster. I considered effects falling below $\boldsymbol{\alpha} < 0.025$ as being statistically significant and commented on non-significant (trend-level) effects for clusters that fell between $\boldsymbol{\alpha} < 0.025$ and $\boldsymbol{\alpha} < 0.1$.

Participants tended to vary in their visual imagery experience and intentionality reports, resulting in unbalanced sample sizes between my contrasts. Consequently, I reran the cluster permutation analysis between intentionality states using a more even sample between spontaneous and deliberate probes, whereby, insofar as a participant reported experiencing more of one intentionality state, a random subsample was extracted to match the frequency of the lesser experienced intentionality state. The results of these subset analyses either revealed no significant differences or corroborated (at trend-level) the differences demonstrated by the full-scale analyses.

5.3. ANALYSIS

5.3.1. Subjective ratings analyses

Statistical analyses of the subsequent behavioural were carried out using R (Version 4.2.3; R Core Team, 2018). All defined linear mixed models (LMMs) were computed using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) R packages. To examine the influence of listening conditions (Ambient, Techno, and Podcast) and time points (Pre-Stress, Post-Stress, and Post-Listening) on self-report of anxiety, I ran separate models for the two stress indices: subjective STAI scores and SCR.

In these models, STAI and SCR levels were the dependent variable, with time point and listening conditions as the fixed effects and an interaction term defined between them, as well as participant included as the random effect.

5.3.2. The relationship between visual imagery and stress-related anxiety reduction

To delineate the unique effects of visual imagery on any reductions of stress indices across listening conditions, I ran linear mixed models each using drops in stress as the dependent variable, with visual imagery incidence and listening condition as the fixed effects, with interaction terms defined between them, and with participant as the random effect. Stress-level reduction was estimated by subtracting measurements of self-reported stress and SCR during the pre-stress period (TP1) from the measurements taken after music listening (TP3). Where interactions between imagery and listening conditions were found, follow-up linear models for each listening condition were estimated in order to explore these effects further.

5.3.3. Effects of individual differences

LMMs were run to assess the influence of general visual imagery and emotion imagery abilities on the frequency of visual imagery probe responses. In three models, each predicting the percentage of either visual imagery, spontaneous imagery, or deliberate imagery, the Vision and Emotion subscales of the Psi-Q were included as fixed effects, and participant and listening condition as random effects.

I next explored whether any effect of visual imagery probes on drops in stress levels depends on general stress reactivity and trait anxiety levels. LMMs were computed to predict each of the stress indices (subjective STAI scores and SCR) as the dependent variable in two separate models, with imagery (all, spontaneous, deliberate) and PSRS and STAI-T scores as fixed effect, and participant and listening condition included as random effects.

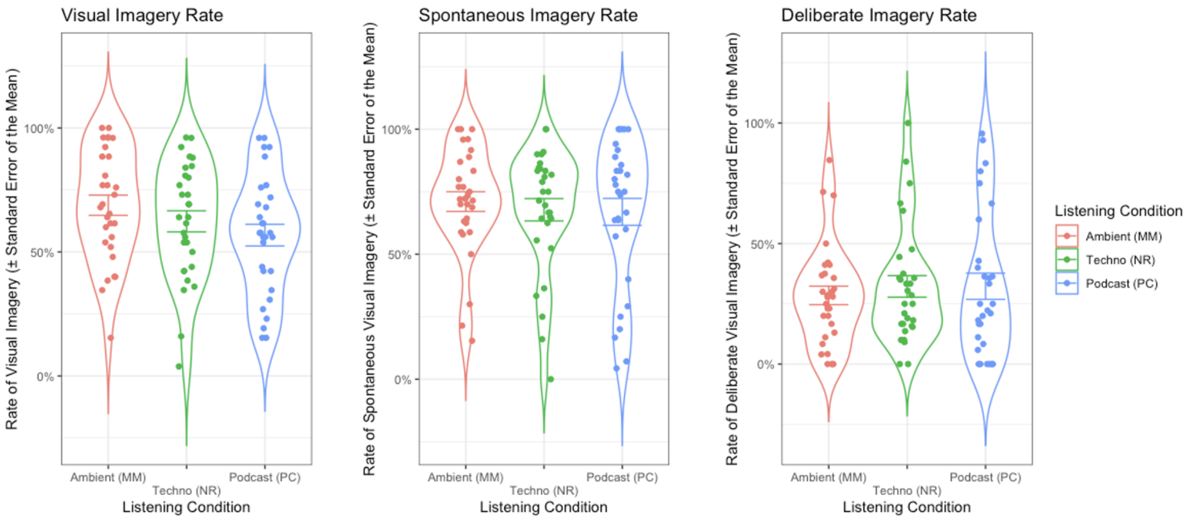
5.4. RESULTS

5.4.1. Incidence rates of visual imagery and its intentionality

Participants reported experiencing visual imagery in response to 62.64% (SD = 48.39) of probes across all listening conditions. Thus, visual imagery was reported more often than not experiencing visual imagery (M = 32.25%, SD = 46.75; note that 5.11%, SD = 22.02, pertain to missed probes). Participants also reported experiencing spontaneous visual imagery (M = 70.42%, SD = 45.66) more often than deliberate visual imagery (M = 29.16%, SD = 45.47). See Figure 5-3 for a visualisation of these rates across the listening conditions.

Figure 5-3

Rates of (spontaneous and deliberate) visual imagery incidence across the three listening conditions (Ambient, Techno, and Podcast tracks).



5.4.2. Neuro-oscillatory characteristics of visual imagery and intentionality

5.4.2.1. Alpha

The cluster-based permutation tests revealed a significant suppression of alpha power between visual imagery and no visual imagery states. This was evident in one cluster in the –2.5 to –0.75 s pre-probe time range, *p* = 0.004, and tended to occupy central and occipital channels most extensively, but also transiently occupied frontal and fronto-central channels (see Figure 5-4). There was also a significant alpha suppression for spontaneous imagery probes (versus no visual imagery probes), in one cluster also in the –2.5 to –0.75 s, *p* = 0.004, in central, left central, bilateral parieto-occipital, and occipital channels (see Figure 5-5). Significant suppression between deliberate and no imagery states was also found in three clusters with a broad dispersion of channels displaying alpha suppression (see Figure 5-6). Cluster 1 (*p* = 0.002, –6.6 to –4.9 s time range) was localised to central as well as fronto-central and occipital channels; in cluster 2 (*p* = 0.008, –10 to –8.9 s time range), alpha suppression was more posterior; while cluster 3 (*p* = 0.022, –8.75 to –8.2 s time range) displayed moments of frontal as well as parieto-occipital suppression.

Trend-level statistical differences between spontaneous and deliberate manifestations of visual imagery were observed, whereby there was greater (right) central, alpha enhancement for the spontaneous imagery condition, *p* = 0.040 (–6.1 to –5.45 s time range).

Figure 5-4

Oscillatory differences between states (visual imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between visual imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Alpha [8-13 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).

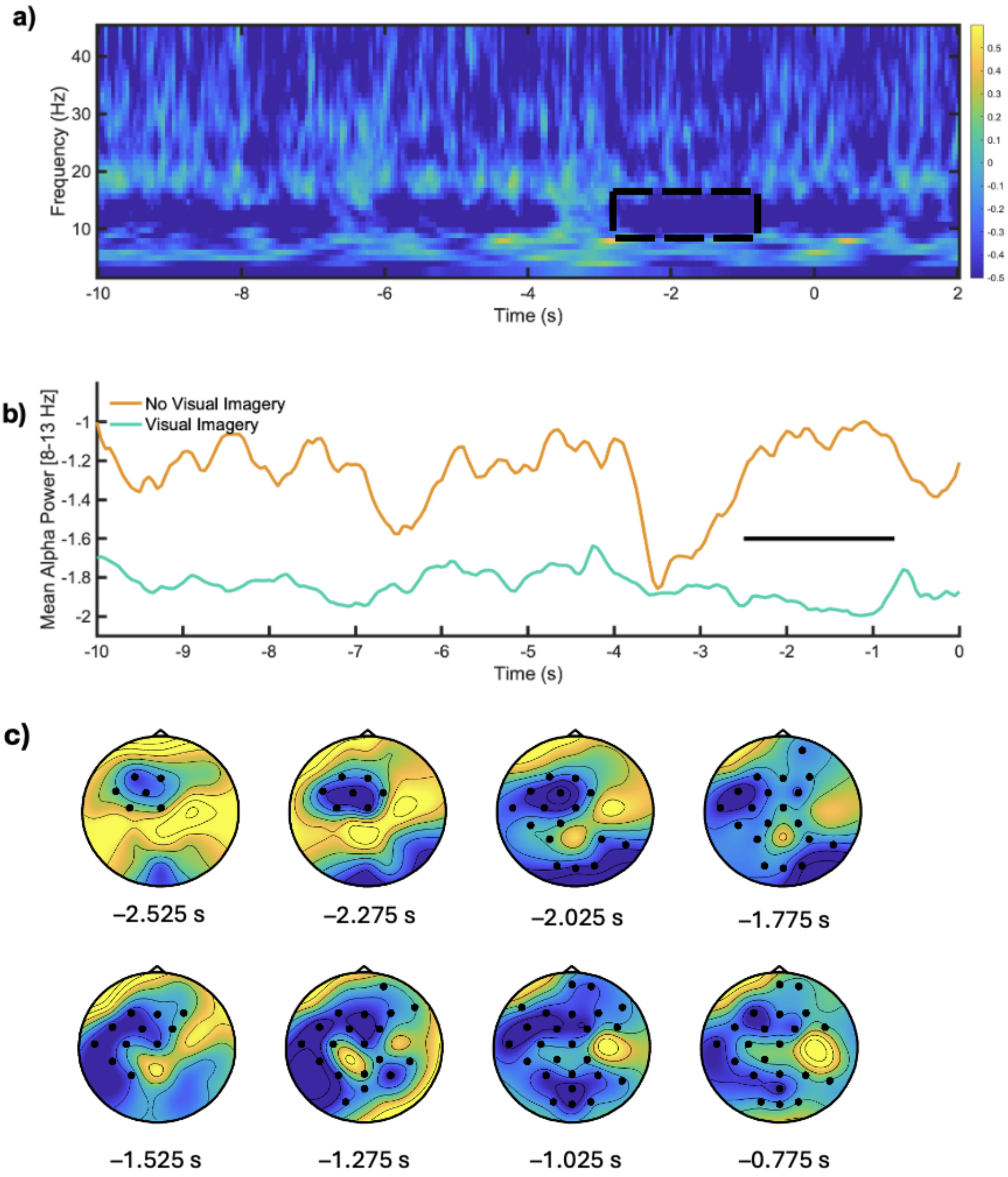


Figure 5-5

Oscillatory differences between states (spontaneous imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between spontaneous imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Alpha [8-13 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).

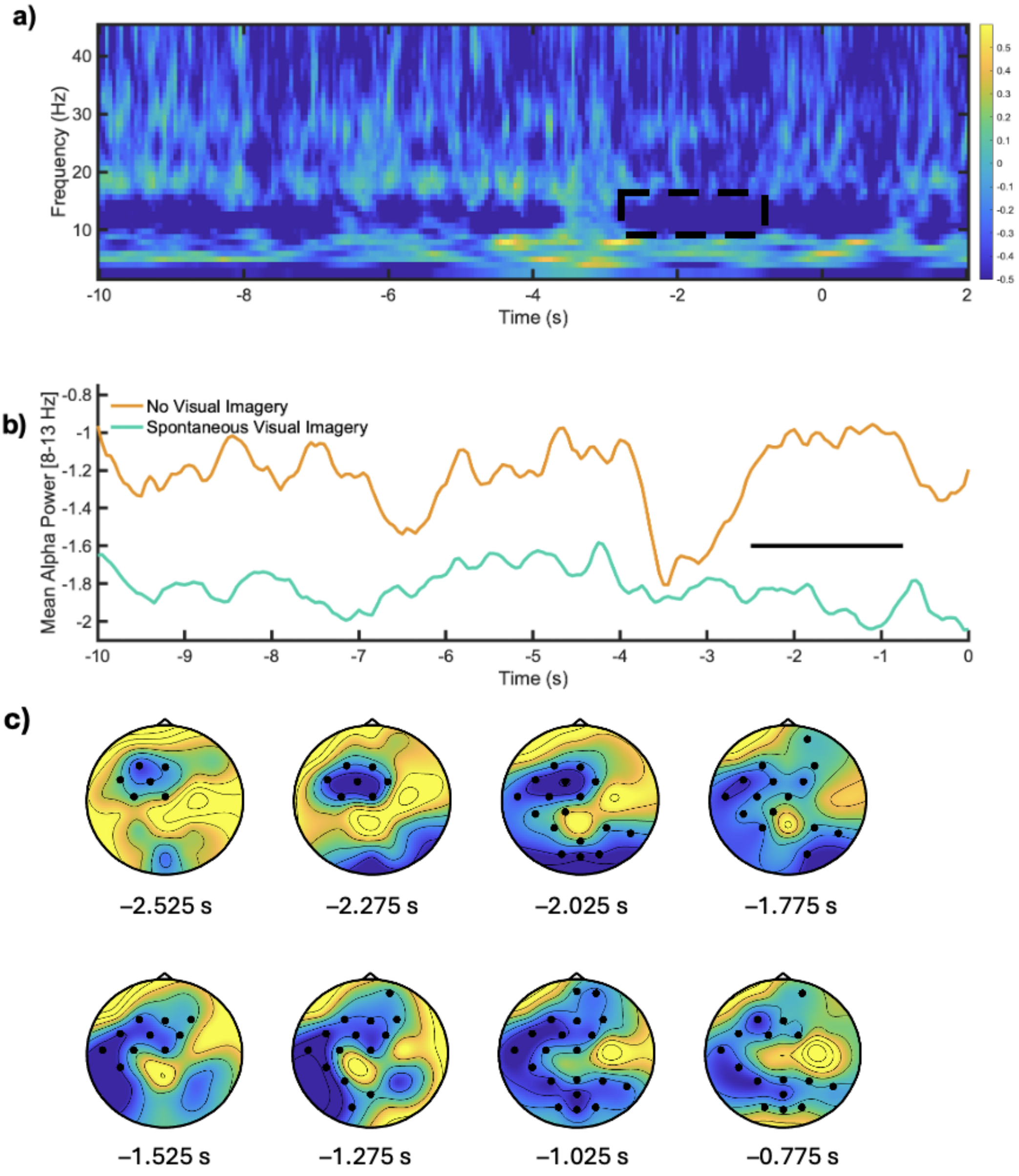
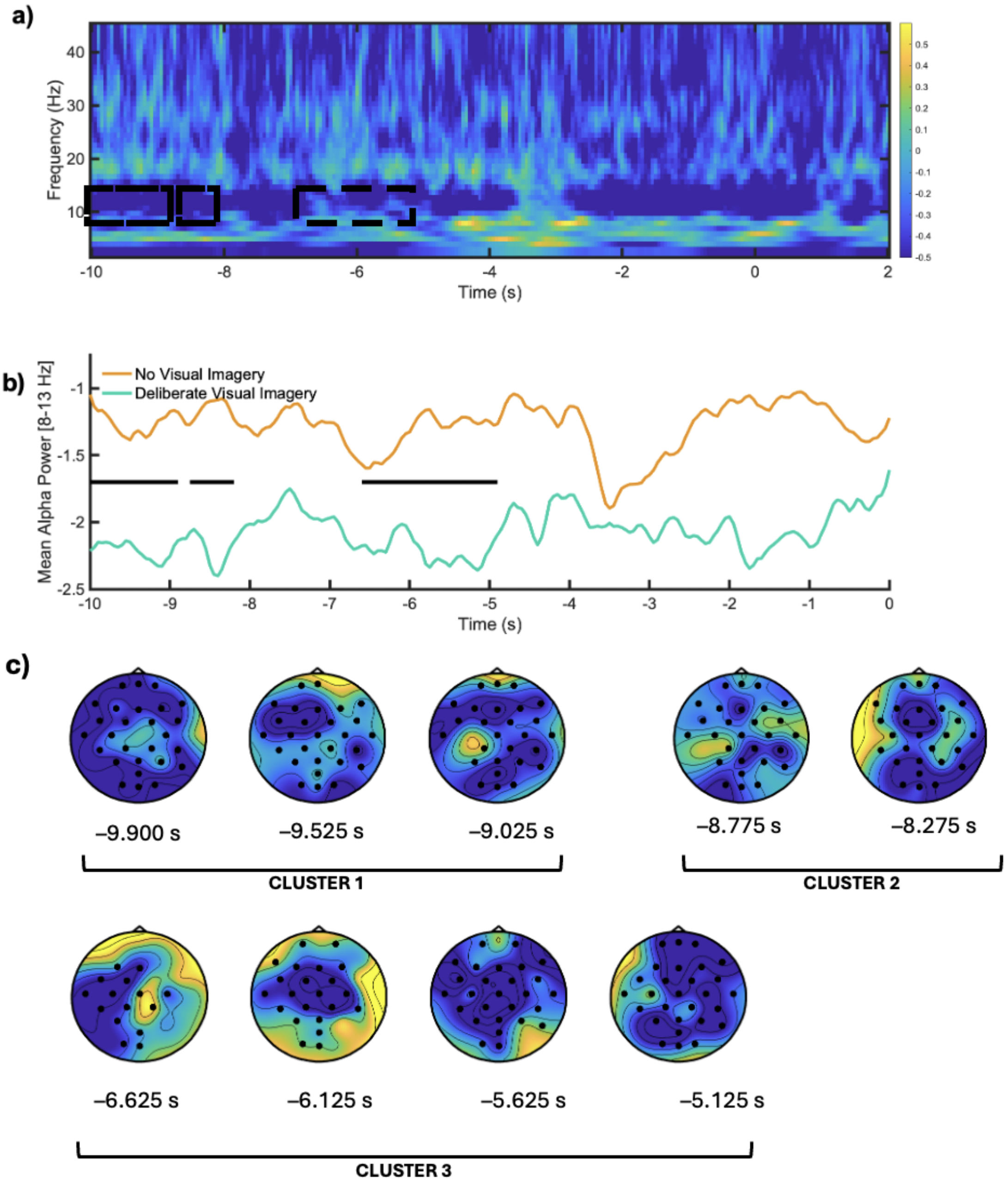


Figure 5-6

Oscillatory differences between states (deliberate imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between deliberate imagery and no visual imagery probes. Significant clusters are indicated by a broken black rectangle. (b) Alpha [8-13 Hz] spectral power averaged over the channels of the clusters. The significant clusters are denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the clusters).



5.4.2.2. Gamma

Statistical differences between visual imagery and no visual imagery probes were found, whereby visual imagery probes were associated with reduced gamma (see Figure 5-7) in six clusters (see Table 5-1). These six clusters were at first located in left central and fronto-central channels earlier before moving to midline/right frontal, bilateral central, and bilateral parieto-occipital channels, and finally becoming (closest to the probe onset) more topographically diffuse but transiently focal in frontal and occipital channels.

Figure 5-7

Oscillatory differences between states (visual imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between visual imagery and no visual imagery probes. Significant clusters are indicated by a broken black rectangle. (b) Gamma [30-45 Hz] spectral power averaged over the channels of the clusters. The significant clusters are denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the clusters).

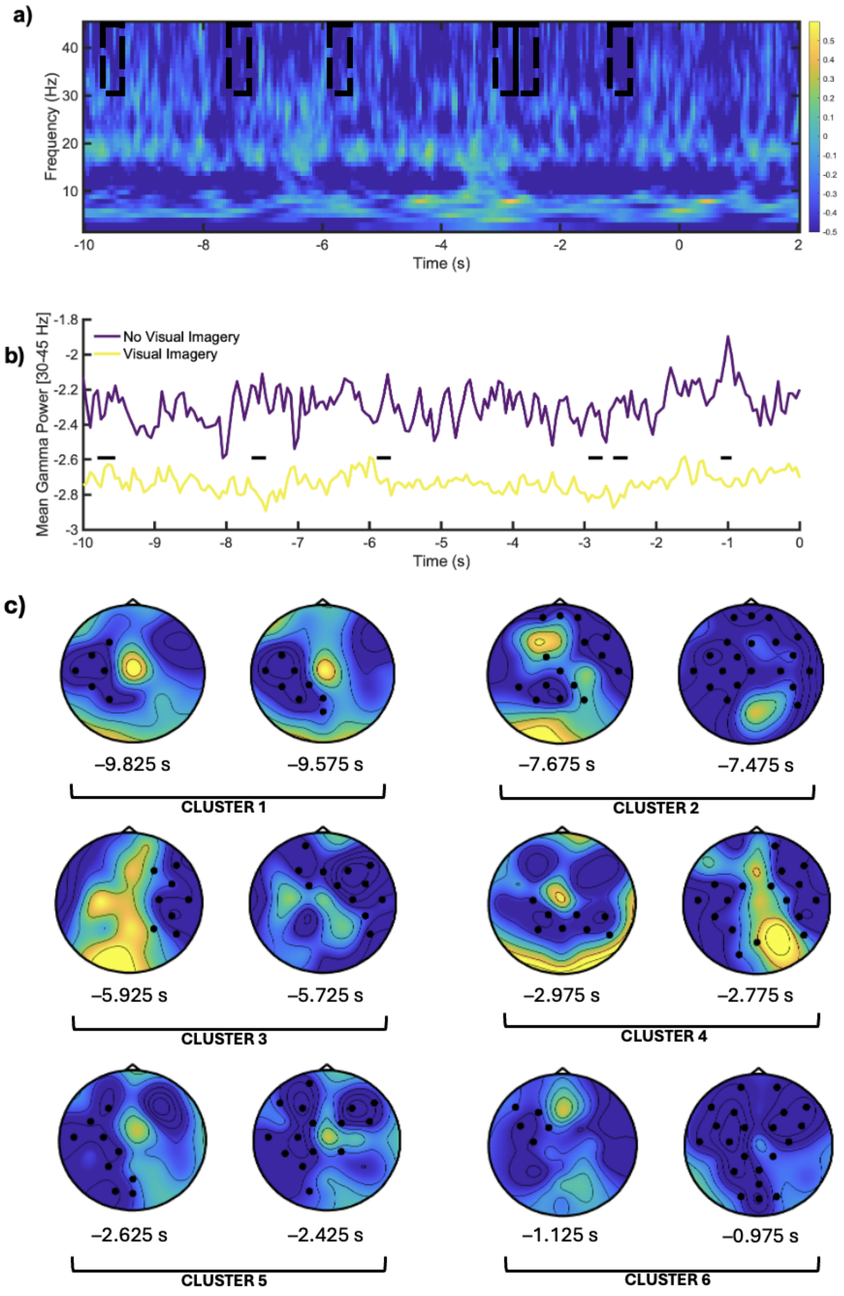


Table 5-1

Cluster statistics representing significant differences in gamma power suppression between the visual imagery and no visual imagery probes, in order of statistical significance.

Clusters displaying significant gamma power suppression	**Time Range (s)**	***p***
Visual Imagery vs. No Visual Imagery		
Cluster 1	–7.65 to –7.45	0.002
Cluster 2	–1.10 to –0.95	0.012
Cluster 3	–5.9 to –5.7	0.016
Cluster 4	–2.95 to –2.75	0.016
Cluster 5	–2.60 to –2.40	0.016
Cluster 6	–9.80 to –9.55	0.016

Gamma suppression was found for both spontaneous (see Figure 5-8) and deliberate (see Figure 5-9) imagery (compared to no imagery probes), with spontaneous imagery gamma suppression present in one right fronto-central and central channel cluster between the –5.9 to –5.7 s pre-probe time range, $p = 0.022$, and deliberate imagery gamma suppression evident in four short-lived clusters. The first cluster ($p = 0.002$, –10 to –9.2 s time range) was diffuse but with focal points in frontal and bilateral parieto-occipital channels while the other three clusters (clusters 2, $p = 0.012$, –7.35 to –7.1 s time range; cluster 3, $p = 0.012$, –8.65 to –8.4 s time range; and cluster 4, $p = 0.018$, –7.60 to –7.45 s time range) largely implicated frontal, fronto-central, and bilateral central electrodes.

Figure 5-8

Oscillatory differences between states (spontaneous imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between spontaneous imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Gamma [30-45 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).

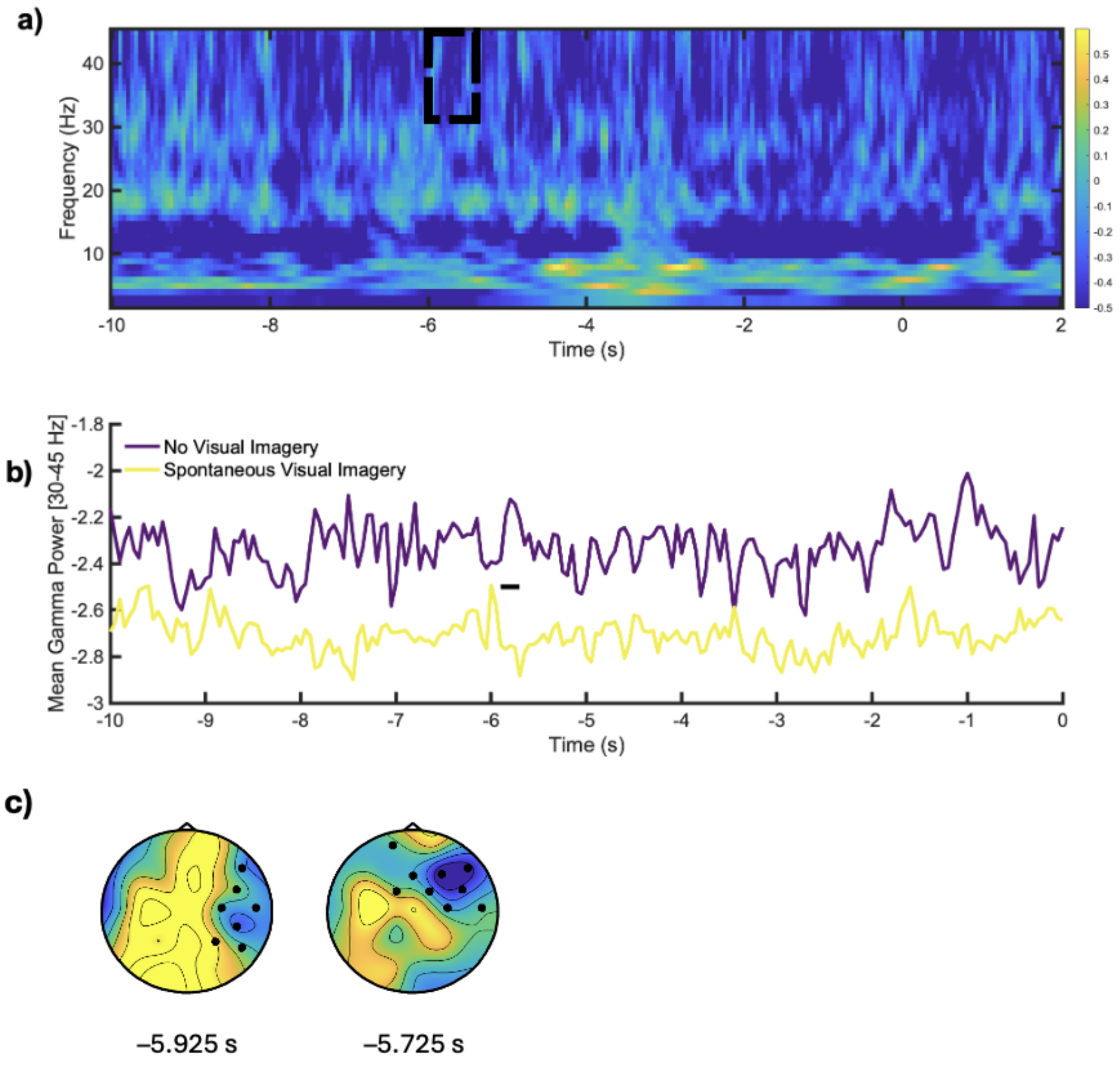
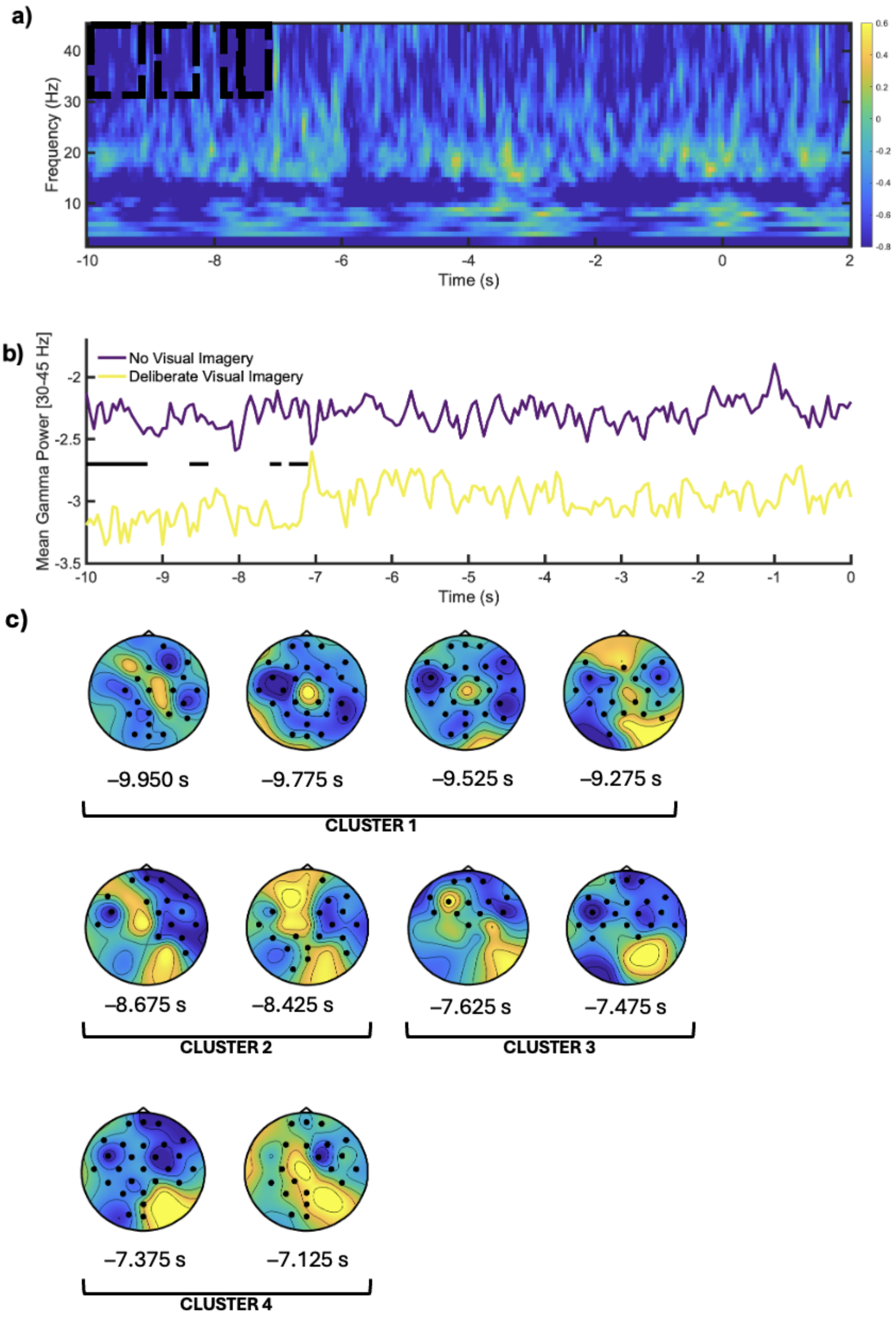


Figure 5-9

Oscillatory differences between states (deliberate imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between deliberate imagery and no visual imagery probes. Significant clusters are indicated by a broken black rectangle. (b) Gamma [30-45 Hz] spectral power averaged over the channels of the clusters. The significant clusters are denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the clusters).



Trend-level statistical differences between spontaneous and deliberate visual imagery revealed greater gamma enhancement in response to spontaneous imagery in one cluster, *p* = 0.050 (–9.75 to –9.70 s time range), that primarily occupied bilateral central, and right parieto-occipital channels.

5.4.2.3. Theta and delta

With regard to theta band, while there were no differences between visual imagery and no imagery probes, and between spontaneous imagery and no imagery probes, deliberate imagery compared to no visual imagery probes was associated with theta suppression, *p* = 0.002 (–2.05 to –1.40 s time range), in midline frontal, fronto-central, central, and occipital electrodes (see Figure 5-10). Further, when comparing spontaneous and deliberate imagery probes, enhanced theta was found in response to spontaneous imagery in four clusters that were largely focal to midline channels (see Figure 5-11). Clusters 1 (*p* = 0.002, –2.25 to –0.75 s time range) and 2 (*p* = 0.006, –9.50 to –8.25 s time range), while slightly more spread, were mainly focal in left and midline fronto-central, left and midline central, midline parieto-occipital, and midline occipital channels. Whereas clusters 3 (*p* = 0.012, –5.55 to –5 s time range) and 4 (*p* = 0.018, –3.70 to –3.10 s time range) were apparent in midline frontal, midline fronto-central, midline central, midline occipital, as well as right temporal and central, and left parieto-occipital sites. In all, neural activation in the theta band mainly concerns the midline channels.

Figure 5-10

Oscillatory differences between states (deliberate imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between deliberate imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Theta [4-7 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).

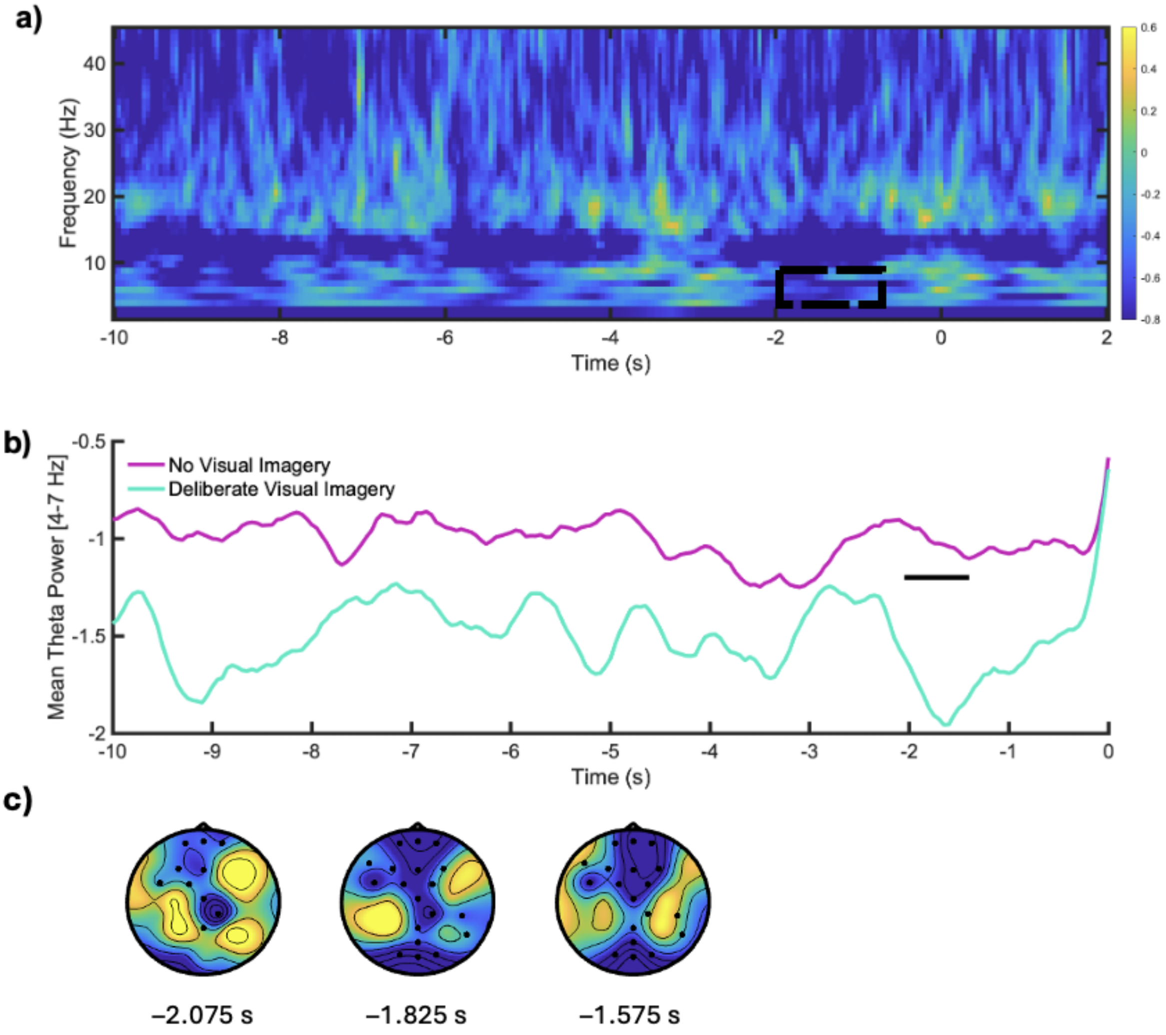
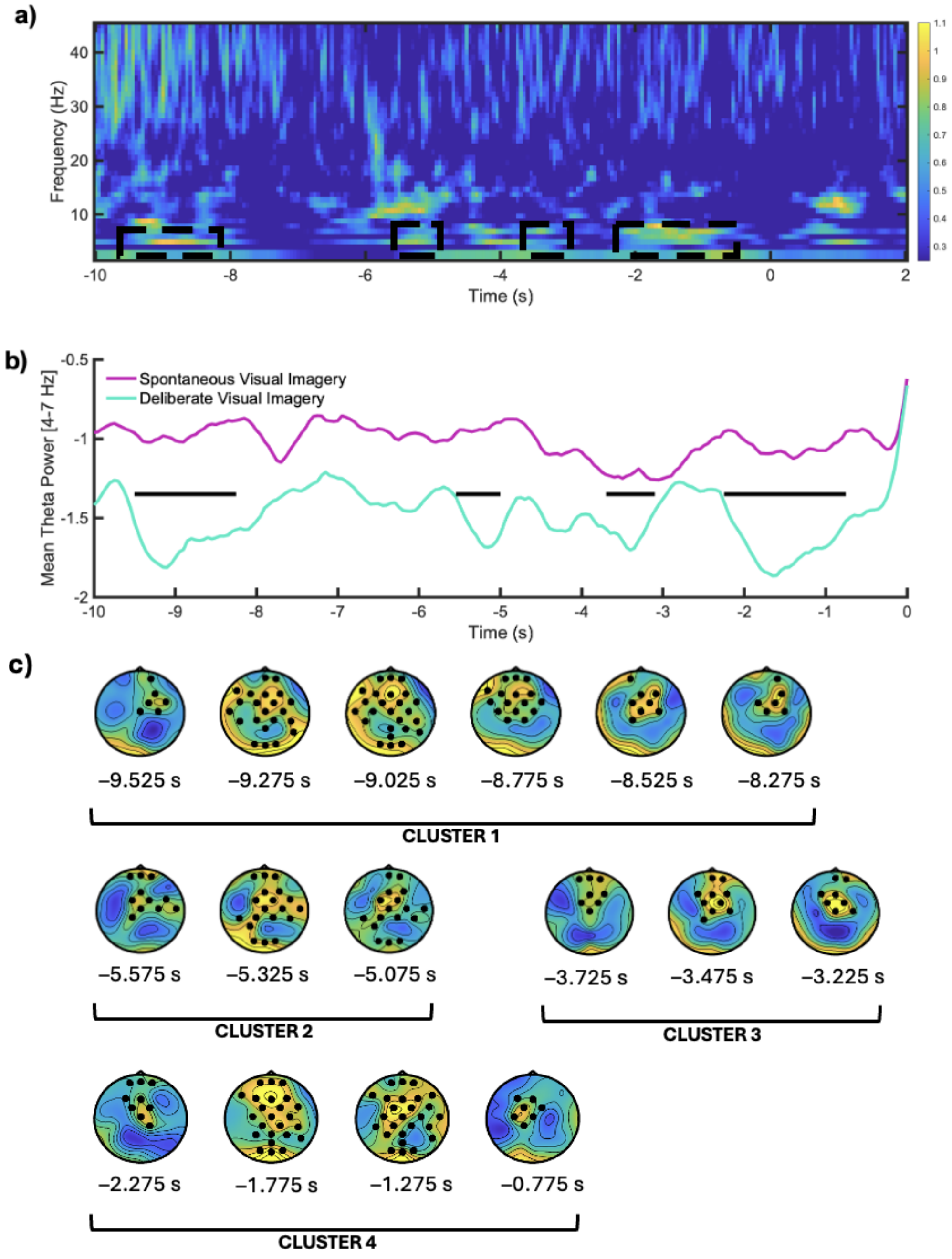


Figure 5-11

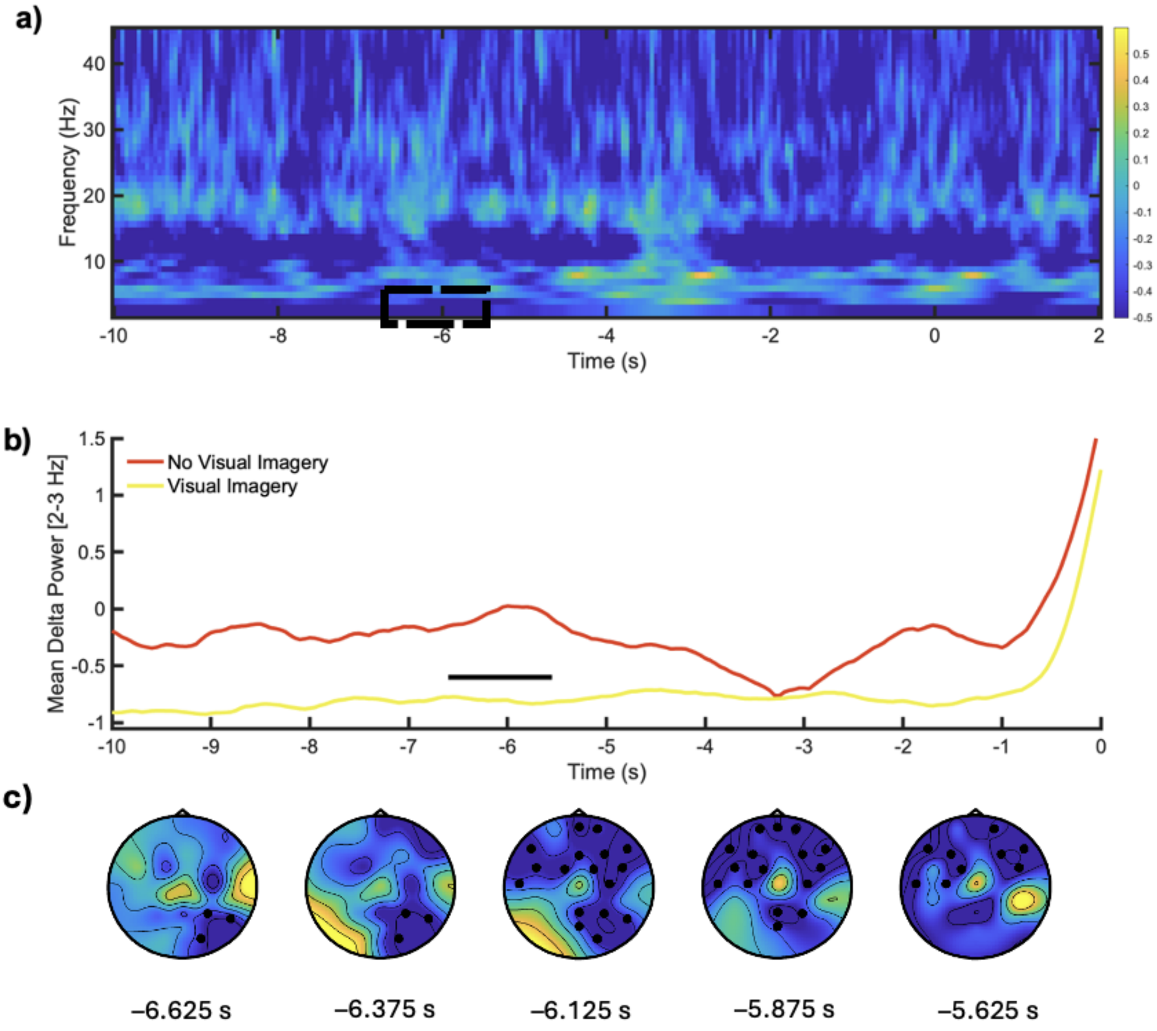
Oscillatory differences between states (spontaneous imagery and deliberate imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between spontaneous imagery and deliberate imagery probes. Significant clusters are indicated by a broken black rectangle. (b) Theta [4-7 Hz] spectral power averaged over the channels of the clusters. The significant clusters are denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the clusters).



Regarding delta, significant suppression was found for visual imagery compared to no visual imagery probes (see Figure 5-12), evident in one cluster, $p = 0.008$ (–6.60 to –5.55 s time range), including channels in bilateral frontal and fronto-central as well as right parieto-occipital areas. No differences were found when comparing spontaneous and deliberate imagery to no imagery probes, though there was trend-level suppression ($p = 0.032$, –6.20 to –5.75 s time range) for spontaneous imagery in midline central and bilateral fronto-central channels, and also for deliberate imagery ($p = 0.046$, –10 to –9.3 s time range) in midline frontal and fronto-central channels. No differences were found between spontaneous and deliberate imagery probes.

Figure 5-12

Oscillatory differences between states (visual imagery and no visual imagery) as a function of time relative to probe onset (0 s). (a) Time-frequency spectrogram averaged across electrode channels of the power difference between visual imagery and no visual imagery probes. Significant cluster is indicated by a broken black rectangle. (b) Delta [2-3 Hz] spectral power averaged over the channels of the cluster. The significant cluster is denoted by the black bar along the x-axis. (c) Topography of the clusters at 250 ms intervals (black markers denote electrodes that were present in the cluster).



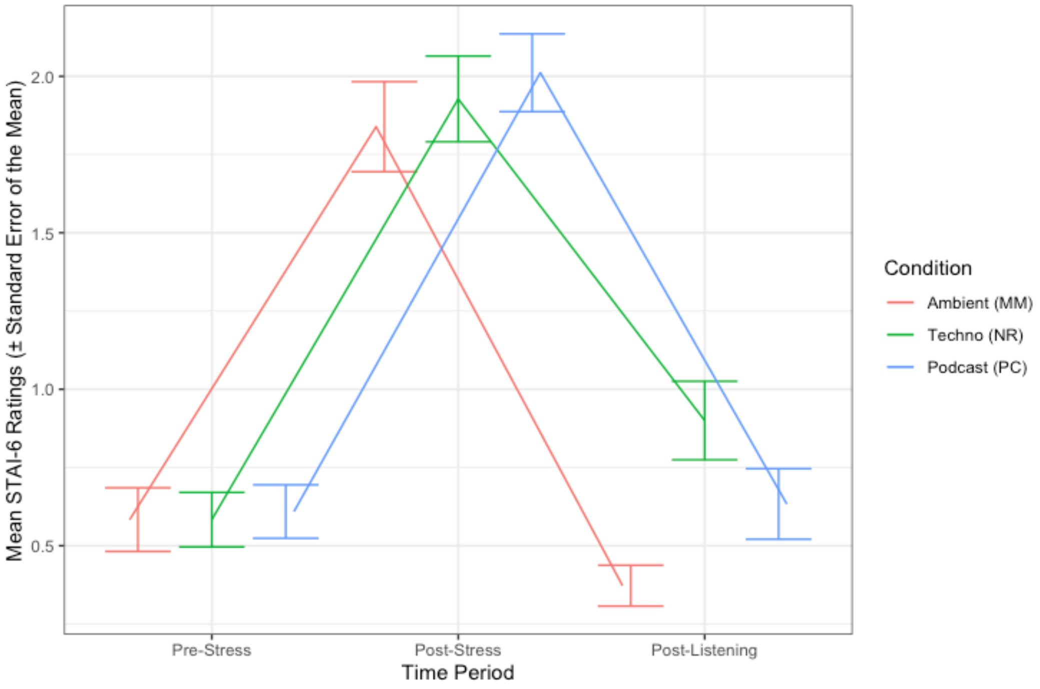
5.4.3. Changes in stress as function of listening condition

The model run as a stress induction manipulation check by examining whether the magnitude of stress induction (i.e., between the pre-stress and post-stress time periods) differed between listening blocks revealed a main effect of time point, $F(1, 142.01) = 444.81, p < 0.001$, but no main effect of listening condition, $F(2, 142.27) = 0.58, p = 0.562$, and no interaction between time point and listening condition, $F(2, 142.01) = 0.49, p = 0.613$. This means that stress was induced in every block to a similar extent.

The model examining how subjectively reported stress-related anxiety levels changed as a function of listening condition and time point revealed a main effect of time point, $F(1, 142.95) = 0.4674, p = 0.493$, a main effect of listening condition, $F(2, 143.37) = 5.87, p = 0.004$, and a significant interaction between time point and listening condition, $F(2, 142.95) = 5.89, p = 0.003$ (see Figure 5-13). Follow-up models were run for each listening condition separately. This revealed a significant difference between baseline (M = 58.33, SD = 55.67) and post-listening (M = 37.22, SD = 35.74) probes for the Ambient track, ß = –0.21, SE = 0.09, $t(29) = –2.44, p = 0.021$, whereby self-reported stress was lower post-listening than at baseline levels. There was also a significant difference between baseline (M = 58.33, SD = 47.7) and post-listening (M = 90, SD = 68.73) for the Techno track, ß = 0.32, SE = 0.12, $t(29) = 2.61, p = 0.014$, whereby after listening to techno music, self-reported stress ratings were higher than after the baseline (pre-stress). Finally, there was no significant difference between baseline (M = 60.92, SD = 45.97) and post-listening (M = 63.33, SD = 60.72) in response to the active control Podcast track, ß = 0.02, SE = 0.12, $t(28) = 0.20, p = 0.844$.

Figure 5-13

Changes in self-reported STAI-6 ratings (± standard error of the mean) over key time periods (pre-stress, post-stress, and post-music) and the three listening conditions (Ambient, Techno, and Podcast).

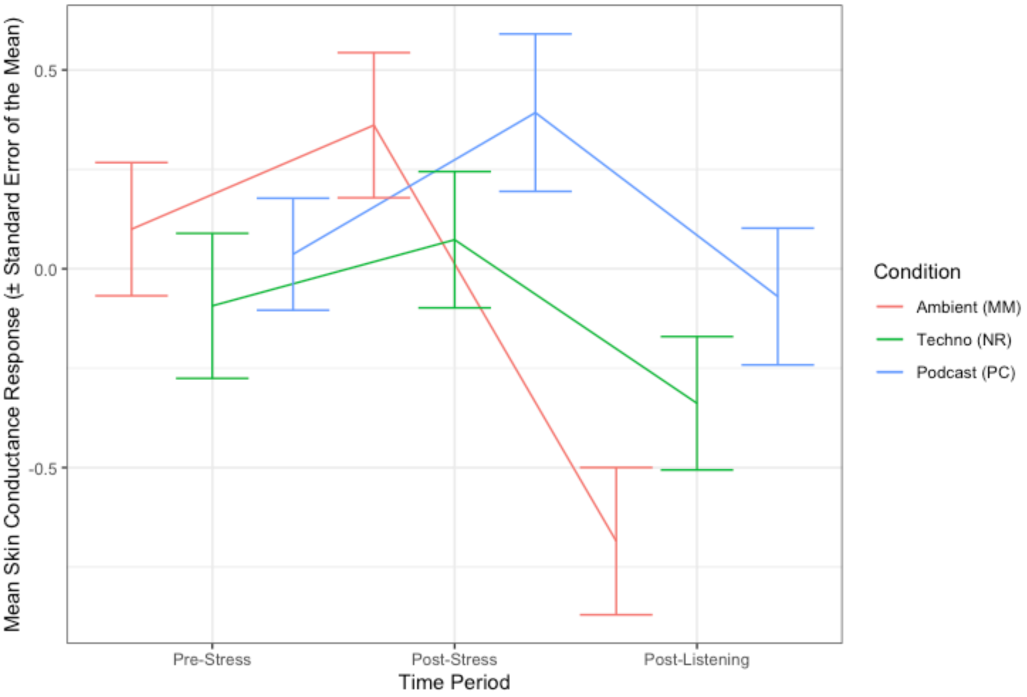


The model assessing the effectiveness of the stress induction manipulation on SCR also revealed there to be a main effect of time period, *F*(1, 133.51) = 5.15, *p* = 0.025, but no main effect of listening condition, *F*(2, 134.20) = 0.64, *p* = 0.640, and no interaction between them, *F*(2, 133.51) = 0.13, *p* = 0.878.

The linear mixed model exploring differences in SCR between time points and listening conditions revealed a main effect of time period, *F*(1, 136.01) = 7.52, *p* = 0.007, but showed no main effect of listening condition, *F*(2, 136) = 1.10, *p* = 0.337, nor any interaction between time period and listening condition, *F*(2, 135.99) = 1.45, *p* = 0.238, despite visualisations of the data displaying a most marked decrease in SCR in response to the Ambient track (see Figure 5-14).

Figure 5-14

Changes in mean (± standard error of the mean) skin conductance response (SCR) over key time periods and across the three listening conditions.



5.4.4. Influence of visual imagery on changes in stress

Having revealed that listening conditions differed in their ability to reduce stress related anxiety, I aimed to examine the potential influence that visual imagery incidence, as distinct from the acoustic features of the different conditions, may play a role in driving stress reduction effects. Indeed, the model assessing the relationship between visual imagery ratings and drops in self-reported stress levels showed there to be a main effect of visual imagery ratings, *F*(1, 83) = 4.34, *p* = 0.040, a main effect of listening condition, *F*(2, 83) = 5.95, *p* = 0.004, and a significant interaction between imagery ratings and listening condition, *F*(2, 83) = 3.63, *p* = 0.031. I ran follow-up linear models of the relationship between rates of imagery and drops in stress levels within each listening condition separately. This revealed a non-significant positive relationship between imagery and drops in stress during the Ambient track, ß = 0.005, SE = 0.004, $t = 1.15$, $p = 0.259$, a non-significant relationship negative during the Podcast track, ß = –0.003, SE = 0.005, $t = –0.48$, $p = 0.636$, but a significant positive relationship during the Techno track, ß = 0.015, SE = 0.005, $t = 3.22$, $p = 0.003$, whereby more imagery was associated with more stress reduction. Assessing these differences with regard to spontaneous and deliberate visual imagery more specifically did not reveal any significant effects of imagery, listening conditions or an interaction suggesting that any potential stress reduction effects of imagery are the same regardless of whether it was spontaneously or deliberately experienced.

Assessment of the relationship between visual imagery ratings in the different listening conditions on drops in SCR showed there to be no significant main effect of visual imagery ratings, $F(1, 67.13) = 3.85$, $p = 0.054$, no significant main effect of listening condition, $F(2, 58.55) = 0.75$, $p = 0.476$, and no significant interaction, $F(2, 59.13) = 0.82$, $p = 0.444$. Similarly, there were no main effects or interactions found when including spontaneous and deliberate visual imagery as fixed effects instead.

5.4.5. Trait attributes of imagery and stress

The model assessing the influence of the Vision and Emotion dimensions of the Psi-Q on visual imagery and deliberate imagery responses during the listening blocks did not reveal any significant associations, although there was a trending relationship between the Vision subscale and spontaneous visual imagery ratings, ß = 5.97, SE = 2.36, $t = 1.75$, $p = 0.091$.

The analyses examining the interaction between (spontaneous and deliberate) visual imagery and trait stress reactivity (PSRS) and anxiety (STAI-T) on the magnitude of drops in subjective stress yielded no significant effects.

Regarding the analyses examining the interaction between (spontaneous and deliberate) visual imagery and trait stress reactivity and anxiety on the magnitude of drops in SCR, the model including visual imagery as fixed effect showed that there was a significant main effect of visual imagery, ß = 0.28, SE = 0.13, $t = 2.18$, $p = 0.033$, only. There were no other main effects or interactions in this model, or the models with spontaneous and deliberate imagery as fixed effects.

5.5. DISCUSSION

Stress and anxiety have emerged as ubiquitous and pressing concerns across populations, associated with a number of mental and physical repercussions. Thus, the development of an accessible means of stress and anxiety management and reduction is an important endeavour for current research. Music presents itself to be an economical, non-invasive, and non-pharmacological method for alleviating stress-related anxiety, and has been evidenced to encourage states of relaxation (Yehuda, 2011) and reductions in subjective and physiological stress-related anxiety and arousal (Chafin et al., 2004; Groarke et al., 2020; Radstaak et al., 2014). However, while previous work has tended to emphasise the role of music's acoustic features in this regard, visual imagery is increasingly purported to be a mechanism by which music is able to do so (Baltazar & Västfjäll, 2020; Panteleeva et al., 2018). The aim of Study 4 was thus to examine the extent to which music-evoked visual imagery acts as a mechanism underlying the stress reduction capacity of music, while also replicating previous studies examining the neuro-oscillatory correlates of music-evoked visual imagery. I hypothesised that music-evoked visual imagery would be (1) associated with posterior alpha suppression, particularly for spontaneous imagery and frontal gamma suppression as demonstrated in previous work; and (2) predictive of drops in levels of physiological arousal (SCR) and self-report ratings (STAI-6) of state anxiety. These hypotheses were possible thanks to recording EEG while administering a probe-caught paradigm that allowed instances of (spontaneous and deliberate) visual imagery to be captured during listening conditions that followed a stress induction task.

5.5.1. Neural data reflects predicted oscillatory characteristics of music-evoked visual imagery

In examining the neuro-oscillatory characteristic of visual imagery, it was possible to confirm my hypothesis and replicate the findings of Study 2 (Chapter 3) and Study 3 (Chapter 4; see Table 5-2 for a summary of all EEG and physiology results). Specifically, findings indicated that music-evoked visual imagery is related to occipital alpha suppression (in replication of Studies 2 and 3) as well as gamma suppression in frontal and central sites (in replication of Study 3). However, the current study was also able to demonstrate that i) deliberate imagery is also significantly associated with posterior alpha suppression (contrary to the previous study's finding that spontaneous imagery is more associated with alpha suppression than deliberate imagery), ii) that spontaneous imagery is associated with enhanced (fronto-central) theta and alpha compared to deliberate imagery; and iii) visual imagery in general is associated with delta suppression in frontal and fronto-central as well as in right parieto-occipital areas.

The finding that deliberate imagery was, like spontaneous imagery, also significantly associated with posterior alpha suppression is interesting and begs the question of why it was not found in my previous study. Here, it is important to once more point out that in the previous study there were no differences in the degree of posterior alpha suppression when comparing spontaneous and deliberate conditions, suggesting that even in that study deliberate imagery may also have been driving alpha suppression. One possibility is that the deliberate imagery experienced in this study, where the music was much more homogenous, was more vivid than what was reported in the previous study, where the musical tracks were changing more frequently.

Table 5-2. Summary of all EEG/physiology studies and key results presented throughout thesis.

Experiment	**Methodology**	**Aims**	**Hypotheses**	**Conditions**	**Outcome Measures**	**Key Findings**
Study 2 (N = 42, Chapter 3)	Retrospective self-report ratings	Examine extent to which static and dynamic music-evoked visual imagery may be associated with differing neural patterns with respect to three regions of interests (frontal, centro-parietal, and parieto-occipital) and three frequency bands of interest (alpha [8–13 Hz], beta [14–30 Hz], and gamma [30–45 Hz]) Explore how imagery-related neural activity differed in key phases of the listening experience (the first and second halves of the piece)	Negative relationship between amount of music-evoked visual imagery and alpha band in parieto-occipital area. Dynamic imagery may involve both beta suppression and enhancement (due to rebound effects) Parieto-occipital gamma enhancement in response to visual imagery generally, but in response to dynamic imagery more specifically	*Imagery Rating Type*: Static Imagery vs. Dynamic Imagery	Self-reports *15-Channel EEG Power*: Alpha (overall, upper and lower), Gamma, Beta	Parieto-occipital alpha **suppression** associated with music-evoked visual imagery Beta **enhancement** to static imagery and **rebound effect** in response to dynamic imagery Gamma **enhancement** to dynamic imagery
Study 3 (N = 30, Chapter 4)	Probe-caught experience sampling	Examine neural correlates of deliberately and spontaneously occurring visual mental imagery in response to music during concurrent EEG recording	Music-evoked visual imagery characterised by alpha power suppression in the posterior regions of the brain Deliberate visual imagery would be less associated with such alpha power suppression (than spontaneous imagery) Spontaneous visual imagery may be associated with increases in theta and delta power	*Listening Condition*: Familiar Relaxing Music Unfamiliar Relaxing Music Familiar Non-Relaxing Music Unfamiliar Non-Relaxing Music	Self-reports *32-Channel EEG Power*: Alpha, Gamma, Theta, Delta	Both visual imagery and spontaneous imagery associated with posterior alpha* and fronto-central gamma **suppression**** Deliberate imagery did not differ from no imagery probes in any frequency band, though also did not show differences from spontaneous imagery in the alpha band No effects found in theta or delta bands
Study 4 (N = 30, Chapter 5)	Probe-caught experience sampling	Examine the role of visual imagery in explaining music's ability to reduce signatures of stress following a stress-inducing task Explore whether spontaneous and deliberate visual imagery would differ in their propensity for stress reduction Examine neural correlates of deliberately and spontaneously occurring visual mental imagery in response to music during concurrent EEG recording	Music with features promoting relaxation will be associated with drops in objective and subjective signatures of stress Incidence of visual imagery will also be shown to be predict the size of these drops in signatures of stress Music-evoked visual imagery will be related to posterior alpha power suppression, with spontaneous imagery showing this activity pattern to a greater extent than deliberate imagery	*Listening Condition*: Ambient, Techno, Podcast (Active Control)	Self-reports *32-Channel EEG Power*: Alpha, Gamma, Theta, Delta Skin Conductance Response (SCR)	Both visual imagery and spontaneous imagery associated with posterior alpha and fronto-central gamma **suppression*** Visual imagery associated with delta **suppression**** Theta **suppression** in response to deliberate imagery** Theta **enhancement** in response to spontaneous imagery** Visual imagery not associated with drops in physiological (SCR) stress.

Note. * replication of previous study's results. ** contrasting with previous study's results.

Another reason for finding deliberate imagery alpha suppression here, but not in Study 3 in the previous chapter, could be due to the nature of the current task. On the one hand, it is possible that participants were more inclined to purposefully conjure visual imagery as a result of the stress induction task. While this could have been done to relax (as was the aim of this study), it is also possible that participants were more struck by the awareness that there was a task to be completed after having taken part in the stress induction task. In other words, coupling a paradigm where high performance was integral to the successful completion of the task (the stress induction task) may have led to the increase in alpha suppression with deliberate imagery, where forming visual imagery during the listening conditions was also perceived to be a task to be successfully achieved (despite emphasis in the verbal instructions for participants to not feel forced to report on any visual imagery if they had not experienced any at all).

The finding that spontaneous imagery was more associated with enhanced (fronto-central) theta and alpha than deliberate imagery is interesting in being in line with findings on unaware mind wandering. Indeed, Polychroni and colleagues (2022) also show evidence for enhanced theta and alpha power during unaware mind wandering, which they suggest to reflect cognitive control of simultaneous thought content (Braboszcz & Delorme, 2011; Klimesch, 1999) and instances in which attention was directed internally (i.e., a decay in perceptual processing; Cooper et al., 2003), respectively. While I predicted these effects but did not see it in Study 3 (in the previous chapter), one possibility is that, again, due to the greater homogeneity of the music heard, participants' spontaneous imagery was often more unaware than during the previous study where music was constantly changing, and where participants may have been focussing on the music more and relatedly largely more aware of their cognition.

The finding that visual imagery was associated with delta suppression is interesting as it differs from the delta enhancements typically reported to mind wandering, and which is held to represent increased focus on and maintenance of internal thought processing and an inhibition of external interference (Harmony, 2013; Harmony et al., 1996; Polychroni et al., 2022). Here, however, it is important to consider the debate regarding whether visual imagery should be considered a form of mind wandering given that, unlike for mind wandering per se, visual imagery during music listening can emerge from attentive listening to music. Indeed, it is possible here that music did not constitute a form of interference, but as a guiding influence on attentive visual thought.

Finally, since delta power has been found to be associated with mind wandering, and since the auditory conditions used in the current study were more homogenous and may have induced more mind wandering, the fact that both forms of visual imagery (spontaneous and deliberate) were associated with delta suppression raises the possibility that delta suppression captures greater focus during imagery episodes than the rest of the listening experience. Indeed, unaware (spontaneous) mind wandering has been shown to be the predominant state in which it is typically experienced and was found to be experienced in the visual domain (Taruffi, 2021). Thus, while I do not attempt to equate visual imagery with mind wandering, the current findings do speak to some degrees of overlap found with regard to their representational format as well as their intentionality state, which should be explicitly compared in future investigations. In any case, this current study, along with Studies 2 and 3 (Chapters 3 and 4, respectively), further contributes to the gradual growth of empirical evidence on the neuroscientific bases of music-evoked visual imagery (Fachner et al., 2019).

5.5.2. Music-evoked visual imagery as a mechanism for reducing subjective stress

Various research has provided subjective and neurophysiological evidence for music's ability to facilitate relaxation in response to a stressful task (e.g., Baltazar et al., 2019; Chafin et al., 2004; Groarke et al., 2020; Helsing et al., 2016; Yehuda, 2011). The results extend that work by demonstrating that the more visual imagery one experiences, the greater the drops in stress they experience in terms of subjective ratings that they report and SCR.

Specifically, the findings show that subjective reports as well as SCR had a stark increase in levels following the stress induction task, followed by a drop in levels after extended listening that was most apparent in response to the relaxing ambient track. Using the magnitude of the drop in stress levels across all the indices (subjective stress ratings and SCR), incidence of visual imagery was shown to only significantly predict drops in the self-report ratings, i.e., it was not shown to be the case with regard to drops in SCR.

This finding may not be so surprising, with other studies also demonstrating divergent results. Some research has found that cardiovascular recovery depends on the genre of the music (Chafin et al., 2004), suggesting that some types of music are more effective in leading to stress recovery than others. Similarly, Groarke et al. (2020) found that music significantly reduced state anxiety when compared to silence (study 1) but not when compared to an active control (study 2). In both cases, music did not lead to reductions in physiological arousal. Further, while Radstaak et al. (2014) did find music to reduce blood pressure and heart rate, this recovery was delayed compared to when in silence. Finally, these findings corroborate older evidence suggesting that music is more likely to excite rather than relax autonomic and muscular activity (Davis & Thaut, 1989). These insights suggest that there is a degree of independence between self-report and physiological outcomes in the case of stress recovery, but that, in all, there is a further notable influence posed by the listening task being presented.

Interestingly, self-report data showed the drop in stress effect to be strongest in response to the techno track. I propose that this finding reflects that the energetic and overwhelming acoustic features of the techno track did not afford as much subjective relaxation, thus allowing the effects of visual imagery to be clearer and more identifiable. In contrast, the acoustic features of the ambient track could have immediately rendered sufficient levels of relaxation, potentially at ceiling levels, to the extent that visual imagery had no additive effect. As such, future investigations should aim to investigate the interactions between visual imagery and specific acoustic features on relaxation more systematically.

A final exploratory interest of this was to examine the extent to which general stress reactivity and trait anxiety would interact with the degree to which visual imagery predicted drops in subjective stress ratings and SCR. These relationships were found to be non-significant, suggesting that these types of inter-individual variability had no influence on the drops in stress nor in the extent to which imagery led to this drop. However, other studies suggest that personality differences may be better determinants of variability in stress recovery to music (Gerra et al., 1998; Ziv et al., 2008).

5.5.3. Implications, limitations and future directions

The current research continues to exhibit the ease and benefits of employing a probe-caught paradigm in assessing difficult-to-measure states such as visual imagery. It presents a move away from more direct questioning that likely evokes deliberate efforts of mental imagery and is one example of a practical methodology for assessing some of the more elusive mechanisms of the music-induced emotion framework (Juslin, 2019; Juslin et al., 2014, 2015).

One limitation to consider is that, given that music typically renders a wide range of multimodal experiences (Juslin & Laukka, 2004), and that imagery is rarely experienced exclusively in the visual domain (see Study 1 in Chapter 2, of this thesis), it is possible that it is not only imagery of visual elements that help to determine reductions in stress levels. Indeed, while Panteleeva et al. (2018) specify mental imagery and autobiographical memories to be potentially beneficial in aiding in negative affect, this could apply to a wider range of domains than just visual imagery in particular. It is possible that other cognitive states, such as mind wandering, could have been present during the extended listening experience and could have more effectively determined reduced levels of stress, as is typically expected during Guided Imagery and Music (GIM) therapy sessions (Dukić, 2022). Nevertheless, visual imagery experienced under explicit and curated instruction has been shown to have a notable role in therapeutic settings (Fachner et al., 2019; Herff et al., 2022, 2023). Thus, while music-evoked visual imagery can be considered beneficial for improving wellbeing during controlled settings, most individuals may not experience it as an automatic strategy with which to alleviate their stress levels.

In contrast, the task of alleviating the stress caused by the different components of the MMST per se may not have played to the strengths of visual imagery. The emotional stressor task (whereby participants listened to a series of unpleasant sounds) may have even rendered more troubling imagery that could have influenced the nature of the visual imagery formed during the listening tasks. Visual imagery has been evidenced to be effective in event simulation and rescripting. This principle is partly what forms the basis of current GIM interventions, whereby a patient is guided by the music in order to give rise to entrenched personal events that emerge in the form of a story through mental imagery (Dukić, 2022; Dukić et al., 2019). Thus, visual imagery may pose a better treatment for types of stress and anxiety that are more contextual and that would more actively benefit from visual imagery, rather than in the current study where it may have been regarded more as a distraction. In the current study, the reductions in stress and the later formation of visual imagery may have therefore acted as separate processes. Nevertheless, visual imagery may still have the ability to provide relief and relaxation to an individual, and future studies should aim to characterise the types of negative experiences that visual imagery, or mental imagery in general, is most effective in aiding.

5.5.4. Conclusion

In sum, I provide first evidence on the potential for music-evoked visual imagery to aid in alleviating feelings of subjective and physiological stress. Neural data was able to confirm that listeners were indeed experiencing visual imagery in response to a relaxing ambient track, a non-relaxing techno track, and a podcast active control track by revealing distinct occipital alpha suppression. Observation of patterns in stress indices indicating that visual imagery experienced in response to a less structurally relaxing track was most effective in reducing experiences of stress provides nuance on the emphasis, to date, of music features as being the primary drivers of relaxation effects in response to music. The following chapter of this thesis follows on in a different direction in exploring emotional responses to music-evoked visual imagery by assessing musical responses and interactions in those with aphantasia, a group of individuals with an inability to form visual imagery.

CHAPTER 6. MUSIC AND APHANTASIA: AFFECTIVE STATES AND EVERYDAY USES

Initial research into aphantasia suggests that aphantasics may experience reduced emotional experiences in response to imagined stimuli. In this two-part online investigation, I sought to explore the emotional experiences of aphantasics within the context of music listening. In Study 5a, I compared 51 aphantasics to 51 control individuals in terms of their experiences of visual imagery, liking, and felt emotional intensity when listening to three film music excerpts. I found significant group differences in terms of visual imagery and felt emotional intensity, but not liking. In Study 5b, I examined aphantasics' ability to recognize emotions conveyed by music, and their patterns of experience of, and engagement with, music in everyday life by comparing the responses of 29 aphantasics with 29 matched controls. No differences in terms of emotion discrimination ability were found. However, aphantasics generally experienced less Reminiscence (dimension from the Adaptive Functions of Music Listening scale) to music, as well as fewer Episodic Memories (dimension from the MecScale). Aphantasics and control listeners did not exhibit differences in terms of sensitivity to musical reward (measured using the BMRQ) or in terms of musical sophistication (measured using the Gold-MSI). Finally, the findings suggest nuanced differences between controls and those with pure and minimal aphantasia. In all, I reveal the influence that aphantasia can have on emotional responses to music and thus provide further evidence for the relationship between visual imagery and music-induced emotion.

6.1. INTRODUCTION

As outlined in Chapter 1, one experience that visual imagery is often associated with is that of emotion induction (Andrade et al., 1997; Holmes & Mathews, 2005; Taruffi et al., 2017) and it has been suggested that visual imagery enhances emotional response through the simulation of imagined situations (Taruffi et al., 2017). While research into the emotional associations of aphantasics is still in its infancy, initial research has found that a lack of visual imagery ability can lead to attenuated emotional response, specifically minimized physiological effects accompanying fear-based visual imagery (Wicken et al., 2021). In fact, some studies show that about half of their aphantasic samples report significantly lower imagery across all other sensory modalities (A. Zeman et al., 2015, 2020), as well as reduced emotional experiences (Dawes et al., 2020). Still, findings in this regard are not always consistent; Zeman et al. (2020) did not find the same lack of emotional experience in aphantasics, instead showing that some aphantasics’ dreams comprise emotions, amongst other experiences such as textual and conceptual forms. Further, aphantasics have even been shown to be subject to similar levels of sensory reactions in response to trauma when compared to controls (Dawes et al., 2020). Thus, it is possible that one would find attenuated emotional reactions in response to the recollections of episodic events in the case of memories (suggesting a constructive role of visual imagery), but not in response to perceptual stimuli more generally (like music).

Links between visual imagery and emotional response are also important in the realm of music (see Taruffi & Küssner, 2019, 2022), and it has become clear that they interact (Balteș & Miu, 2014; Day & Thompson, 2019; Hashim et al., 2020). For example, a study by Cespedes-Guevara and Dibben (2022) found links between the narrative and content of listeners’ visual imagery reports and emotional descriptions that had been provided prior to music listening. In line with Juslin and Västfjäll's (2008) proposal that visual imagery is one of eight mechanisms (Juslin, 2013) potentially mediating the relationship between music and emotional response, a qualitative study into the content of visual imagery showed that emotional experiences occupied a second-order theme amongst other prevalent subthemes (e.g., abstract imagery, memories; see Study 1 in Chapter 2). Further, Vuoskoski and Eerola (2015) argued that visual imagery evoked by narrative descriptions of a sad music track may have enhanced feelings of sadness indicated via a bias in word memory recall. Moreover, in a study by Hashim et al. (2020), which aimed to examine the effects of suppressed visual imagery on emotional response, it was found that listeners required to do an eye-movement task experienced significantly lower amounts of (and less vivid) visual imagery as well as minor reductions in emotional response when compared to a no-task control condition. In all, it would appear that visual imagery experience and emotional response are strongly connected, and that aphantasics may demonstrate a reduced ability to experience emotions from music.

Finally, a pressing question is how reduced visual imagery ability may affect the aesthetic judgement of music. The vividness of the visual imagery that music induces has previously been shown to be a strong predictor of the aesthetic pleasure of poetry: specifically, haikus and sonnets (Belfi et al., 2017). Examining the factors that might predict music's aesthetic appeal, Belfi (2019) similarly found that the vividness of visual imagery followed only felt emotional valence in explaining the aesthetic appeal of music from classical, jazz, and electronica genres. One possibility, therefore, is that aphantasics, due to their reduced visual imagery ability, may show reduced aesthetic pleasure and reward from music listening, and accordingly, may show reduced engagement with it in everyday life.

6.2. THE PRESENT CHAPTER

Taken together, existing literature supports the idea that a lack of visual imagery could be associated with reduced, or dampened, emotional response and aesthetic appeal in relation to music. Where previous studies have approached the question of the relationship between music-evoked visual imagery and emotion induction by measuring reaction times of visual imagery and emotion onset (Day & Thompson, 2019), by using emotive written descriptions with the potential to enhance visual imagery (Vuoskoski & Eerola, 2015), and by active suppression of visual imagery (Hashim et al., 2020) amongst the general population, my research sought to contribute to this question by comparing the experiences of aphantasics with the experience of a sample of typical listeners.

The aim of the research presented in this chapter was to understand the impact that having little-to-no visual imagery ability may have on the emotional experience of music (Study 5a). Further, I aimed to understand whether this may lead to distinct patterns in emotion recognition ability as well as in the experience and engagement with music in everyday life (Study 5b). Taken together, by providing a general impression of aphantasics' affective and aesthetic experiences of music, I hoped to both guide further research into the understanding of aphantasia as a condition within the population, and to advance understanding of visual imagery during music listening.

6.3. STUDY 5a: comparing the emotional and aesthetic responses to music of aphantasic and general listeners

Aphantasics have been proposed to show weaker affective experiences across a range of imagined contexts and situations (Dawes et al., 2020; Kay et al., 2022; Wicken et al., 2021; A. Zeman et al., 2015, 2020), but is this also true of music-related experiences? In survey 1, I sought to gain a general understanding of the differences between aphantasic and control listeners' affective and aesthetic responses to music. I recruited a sample of listeners with aphantasia and compared them to a subset of age- and gender-matched typical listeners previously recruited in Study 1 (Chapter 2). Participants listened to the same three short film excerpts (Eerola & Vuoskoski, 2011) used in Study 1 (in Chapter 1) and, after each excerpt, were asked to rate their experiences of visual imagery prevalence and vividness, music liking, and felt emotional intensity in response using individual continuous Likert scales.

I predicted that I would observe significantly lower reports of visual imagery prevalence (how much imagery was experienced) and visual imagery vividness (how clearly it was imagined) in aphantasics than in the control group (Hypothesis 1). In addition, due to past findings linking a lack of visual imagery ability to a dampening of emotional responses to imagined stimuli (Wicken et al., 2021), it was predicted that the aphantasic listeners would provide lower felt emotional intensity and liking ratings towards the film music excerpts than the control group (Hypothesis 2). In line with the idea that vividness of visual imagery is strongly linked to emotional responses to music (Hashim et al., 2020; Vuoskoski & Eerola, 2015) and a strong predictor of aesthetic appeal (Belfi, 2019), I further predicted that I would find positive relationships between visual imagery prevalence and vividness ratings provided by the aphantasic and control listeners and ratings of liking and felt emotional intensity in response to the film music excerpts (Hypothesis 3). Finally, I aimed to explore associations between music-evoked visual imagery prevalence and general imagery ability (tested using the VVIQ). I predicted that imagery prevalence would be positively related to VVIQ mean scores (Hypothesis 4).

6.4. METHODS

6.4.1. Participants

76 complete submissions were initially received from a pool of self-proclaimed aphantasics. These aphantasics were recruited through dedicated Facebook groups.

From these, I set a criterion to include only *pure* (individuals that report seeing no image at all in their mind's eye in response to all VVIQ items, and therefore obtain the lowest summed VVIQ score of 16 out of 80) and *minimal* aphantasics (individuals that report seeing very few images in their mind's eye in response to VVIQ items, and therefore obtain a summed score of 17–30 out of 80). Consequently, 51 participants remained (*pure* = 33, *minimal* = 18) and were included in further analyses (aged 21–83, 39 females, 11 males, 1 prefer not to say, M = 53.6, SD = 14.8). In terms of location, 39.2% (n = 20) of respondents were residents of the UK, 33.3% (n = 17) of the USA, 11.8% (n = 6) of Canada, 5.9% (n = 3) of Australia, 2.0% (n = 1) of Germany, 2.0% (n = 1) of France, 2.0% (n = 1) of Luxembourg, 2.0% (n = 1) of Taiwan, and 2.0% (n = 1) of Indonesia.

For the control sample, I included a subset of the sample recruited using Prolific (an online platform that facilitates participant recruitment; www.prolific.com) in Study 1 (Chapter 2; aged 18–66 years, 153 females, 198 males, 2 prefer not to say; M = 26.41, SD = 9.41), with the aim of matching them as closely as possible in terms of age and gender. Visual inspection of the original dataset collected in Study 1 (Chapter 2) showed a prominent positive skew in age. An initial attempt to match the control sample with the aphantasic sample as closely as possible led to the exclusion of 13 respondents from both groups at the 67–83 age range. Thus, to avoid these exclusions, I recruited a further 13 general respondents within that age range to facilitate age and gender matching, which resulted in 51 control respondents successfully closely matched in age and gender (aged 21–77 years, 39 females, 11 males, 1 prefer not to say, M = 50.0, SD = 15.5) that were included in further analyses. All individuals within my control sample possessed summed VVIQ scores of 30+, making them an appropriate non-aphantasic control group due to their sufficient general visual imagery abilities. In terms of location, 35.3% (n = 18) of respondents were residents of the UK, 15.7% (n = 8) of Canada, 7.8% (n = 4) of Germany, 5.9% (n = 3) of the USA, 5.9% (n = 3) of Portugal, 5.9% (n = 3) of Poland, 3.9% (n = 2) of Chile, 3.9% (n = 2) of Latvia, 3.9% (n = 2) of Spain, 2.0% (n = 1) of Italy, 2.0% (n = 1) of Greece, 2.0% (n = 1) of Israel, 2.0% (n = 1) of the Netherlands, 2.0% (n = 1) of Mexico, and 2.0% (n = 1) of Sweden.

Independent samples t-tests confirmed that there were no differences between the two participant groups in terms of age, $t(99.7) = 1.21$, $p = 0.230$, but that there was a significant difference between the aphantasic (M = 1.17, SD = 0.28) and control groups (M = 3.68, SD = 0.75) in terms of their VVIQ mean scores, $t(63.6) = 22.29$, $p < 0.001$.

6.4.2. Ethics Statement

This survey has received ethical approval from the Ethics Committee of the Department of Psychology at Goldsmiths, University of London. Participants provided online written consent for the collection, analysis, and potential publication of their data in a scientific journal. They also received monetary compensation for their time.

6.4.3. Materials and stimuli

Three film music stimuli conveying happy, tender, and fearful emotions were selected from Eerola and Vuoskoski's (2011) database. These excerpts were obtained from the catalogue of extended 1-min film excerpts (see Appendix of Vuoskoski et al., 2012), validated to be unfamiliar to most listeners. The excerpt conveying happy emotions was taken from The Untouchables soundtrack (track 6, number 071 from Eerola and Vuoskoski's set of 110 tracks). The tender excerpt is from the Shine soundtrack (track 10, number 042 from the 110 set). Finally, the fearful excerpt is from the Batman Returns soundtrack (track 5, number 011 from the 110 set). These excerpts were chosen in Study 1 (in Chapter 2) to reflect only a subset of contrasting emotions that could induce a rich and varied range of visual imagery content. The main purpose of that study was not to explore music-induced emotions; thus, it did not include any other emotions from Eerola and Vuoskoski's database (namely, anger and sadness). To ensure uniformity amongst the musical excerpts, as well as to control the overall length of the survey, all excerpts were edited to last a duration of 45 s using Audacity (Version 2.3.2.0). I did not anticipate this to compromise the excerpts' cited abilities to induce their intended emotions, given that the original set of 110 tracks from Eerola and Vuoskoski's collection had included even shorter segments of these same excerpts that maintained their validity. The excerpts were also edited to finish with a fade-out to avoid an abrupt ending.

Visual imagery ratings were obtained using two items from Pekala's (1991) Phenomenology of Consciousness Inventory (PCI), a 53-item questionnaire that measures a variety of personal perceptual experiences revolving around consciousness. Participants were asked about the prevalence of imagery in their experience (1 = I experienced no visual imagery at all, to 7 = I experienced a great deal of visual imagery [Q.12]), and the vividness of their imagery (1 = My visual imagery was so vague and diffuse, it was hard to get an image of anything, to 7 = My visual imagery was so vivid and three-dimensional, it seemed real [Q.18]).

Liking ratings in response to the music were measured using a 5-point Likert scale, from 1 = Dislike a great deal, to 5 = Like a great deal. Felt emotional intensity ratings were also measured using a 5-point Likert scale, from 1 = Not at all intense to 5 = Extremely intense.

Finally, the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) was administered as an independent measurement of visual imagery ability. The VVIQ comprises 16 statements to which individuals are instructed to form a visual mental image in their minds (e.g., 'Visualize a rising sun. Consider carefully the picture that comes before your mind's eye.', 'A rainbow appears', etc.). The participants rated the vividness of their visual mental image of the statement along a 5-point Likert scale from 1 = No image at all, you only 'know' that you are thinking of an object to 5 = Perfectly clear and vivid as real seeing. VVIQ possesses a Cronbach's alpha reliability score of $\boldsymbol{\alpha} = 0.74$.

6.4.4. Procedure

Participants were first provided with the aims, instructions, and a definition of visual imagery as "*the spontaneous formation of visual images or pictures in your mind's eye. Your imagery experience is completely subjective, and it is completely acceptable not to have experienced any imagery at all.*" The online survey took approximately 12 mins to complete and was built and presented using Qualtrics. Aphantasic participants were invited to take part in the survey via email, whereas the control sample accessed the link through an advertisement distributed on Prolific. The presentation order of musical stimuli was randomised across participants.

Participants were advised to use good-quality, preferably noise-cancelling headphones (though it was not formally checked whether this advice was adhered to) and to have minimal outside disturbance throughout the study. For the three main trials, participants were first presented with the film music excerpt. They were instructed to listen to the whole excerpt before they would be presented with questions regarding their experience of the music on the following page. After listening, participants were asked to rate the amount and vividness of their visual imagery. Participants were also asked to rate how much they liked the music, and how intense any felt emotional response to the music was. At the end of the survey, participants completed the VVIQ.

6.5. ANALYSIS

Statistical analyses were run in R (Version 4.2.3; R Core Team, 2018). To account for likely violations in normally distributed data in at least one of the sample groups, I opted to run linear mixed effects models to assess differences in the relationships between the aphantasic and control groups in terms of their experiences to the film music excerpts. To this end, four models were run, using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages in R, with the four behavioural measures (prevalence and vividness of visual imagery, liking, and felt emotional intensity) as dependent variables, with Group (aphantasic and control) as fixed effect as well as participant and musical excerpt as random effects in each model.

In an additional set of analyses, I sought to check whether there were any qualitative differences between those within the aphantasia sample who were either pure aphantasics (individuals who provided the lowest VVIQ scores, i.e., experiencing no visual imagery at all) or minimal aphantasics (individuals who provided VVIQ scores of up to 30, i.e., experiencing minimal amounts of visual imagery) by rerunning the same models to assess these subgroups against controls separately. Independent samples t-tests maintained that there were negligible differences in age between the pure aphantasics (M = 53.88, SD = 14.11) and controls (M = 49.98, SD = 15.41), $t(222.25) = 2.07$, $p = 0.040$, and no differences between the minimal aphantasics (M = 53.11, SD = 15.83) and controls (M = 49.98, SD = 15.41), $t(90.8) = 1.26$,

$p = 0.212$. Bonferroni correction was applied to these models to correct for multiple comparisons, leading to an adjusted alpha level of 0.013.

Further, I ran two additional linear mixed models to address my hypotheses regarding the relationship between the prevalence and vividness of visual imagery and the liking, felt emotional intensity, and VVIQ mean scores. Prevalence and vividness were entered as dependent variables in each of the two models, with liking, felt emotional intensity, and VVIQ entered as fixed effects, including participant, musical excerpts, and sample group as random effects. Bonferroni correction was also applied here, leading to an adjusted alpha level of 0.025.

6.6. RESULTS

6.6.1. Prevalence

Out of all the ratings of the 51 pure and minimal aphantasic respondents collapsed across the three music excerpts, 5.9% (n = 3) reported experiencing at least mild levels (a rating of 2 and upwards) of visual imagery during listening, while 94.1% (n = 48) reported experiencing no visual imagery at all. Similarly, regarding visual imagery vividness, 5.9% (n = 3) participants also reported at least mild levels of vividness (a rating of 2 and upwards).

Out of all the ratings of the 51 control respondents collapsed across the three music excerpts, 94.1% (n = 48) reported experiencing at least mild levels (a rating of 2 and upwards) of visual imagery during listening, and 5.9% (n = 3) reported experiencing no visual imagery at all. Regarding visual imagery vividness, 90.2% (n = 46) participants also reported at least mild levels of vividness (a rating of 2 and upwards).

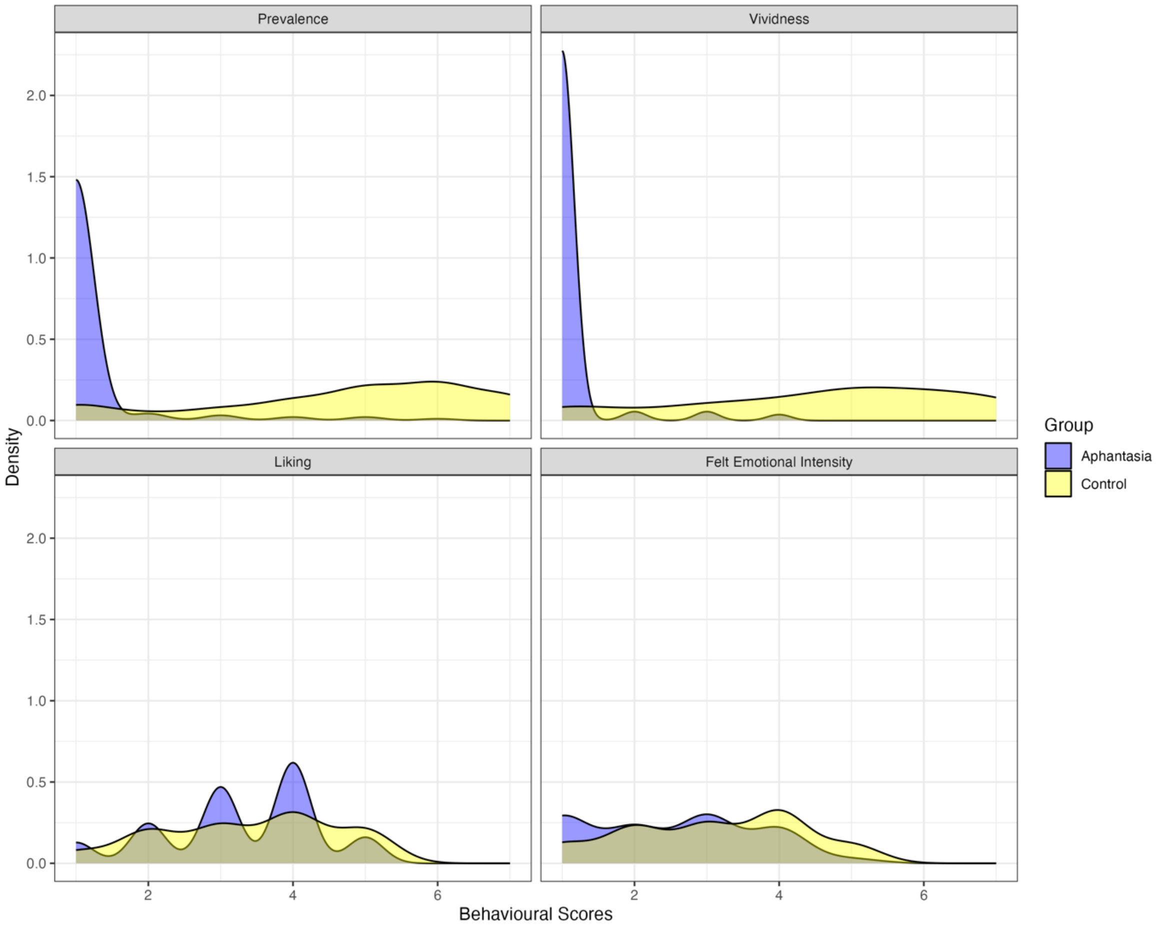
6.6.2. Differences in behavioural data between aphantasic and control groups

In the model predicting prevalence of visual imagery, results indicated significant differences between the aphantasic (M = 1.19, SD = 0.75) and control groups (M = 4.70, SD = 1.87), whereby control participants experienced more visual imagery than aphantasics (ß = 3.49, SE = 0.24, $t = 14.57$, $p < 0.001$, partial η2 = .68, 95% CI [0.60, 1.00]). The model predicting vividness of visual imagery also showed there to be significant differences between the aphantasic (M = 1.11, SD = 0.49) and control groups (M = 4.53, SD = 1.89), with controls reporting higher vividness (ß = 3.41, SE = 0.26, $t = 13.21$, $p < 0.001$, partial η2 = .65, 95% CI [0.56, 1.00]). The model predicting felt emotional intensity similarly showed the control group

(M = 3.08, SD = 1.21) to report significantly higher felt emotional intensity in response to the music than aphantasics (M = 2.49, SD = 1.18; ß = 0.58, SE = 0.19, *t* = 3.04, *p* = 0.003, partial η2 = .09, 95% CI [0.02, 1.00]). However, the model predicting music liking did not exhibit any differences between the aphantasic (M = 3.27, SD = 1.09) and control (M = 3.36, SD = 1.22) groups (ß = 0.09, SE = 0.13, *t* = 0.69, *p* = 0.490). Assessment of model assumptions using graphical means (histograms and Q-Q plots) indicated that residuals and random effects for both models satisfy normality. See Figure 6-1 for a visualisation of these results.

Figure 6-1

Density plots of the four behavioural measures (prevalence and vividness of visual imagery, music liking, and felt emotional intensity) between the aphantasic and control groups.



I reran these models to compare differences between controls and pure as well as minimal aphantasics. Please find detailed results and visualisations in the Appendix D (Table D.1 and Figures D.1 and D.2). In the model comparing prevalence of visual imagery between pure aphantasics and controls, results indicated significant differences between the pure aphantasics (M = 1.11, SD = 0.76) and control groups (M = 4.70, SD = 1.87), whereby control participants experienced more visual imagery than pure aphantasics. The model comparing prevalence of visual imagery between minimal aphantasics and controls also indicated significant differences between the minimal aphantasics (M = 1.33, SD = 0.87) and control groups (M = 4.70, SD = 1.87), whereby control participants experienced more visual imagery than minimal aphantasics.

The model predicting vividness further showed there to be significant differences between the pure aphantasics (M = 1.00, SD = 0.00) and control groups (M = 4.53, SD = 1.90), with controls reporting higher vividness. There were also differences in terms of vividness between minimal aphantasics (M = 1.27, SD = 0.74) and controls (M = 4.53, SD = 1.90), whereby controls reported higher vividness.

In terms of felt emotional intensity, the models similarly showed the control group (M = 3.08, SD = 1.21) to report significantly higher felt emotional intensity in response to the music than pure aphantasics (M = 2.61, SD = 1.20) as well as minimal aphantasics (M = 2.26, SD = 1.11).

However, the models predicting music liking did not exhibit any differences between controls (M = 3.36, SD = 1.22) and either the pure (M = 3.32, SD = 1.12) or minimal (M = 3.19, SD = 1.03) aphantasic groups.

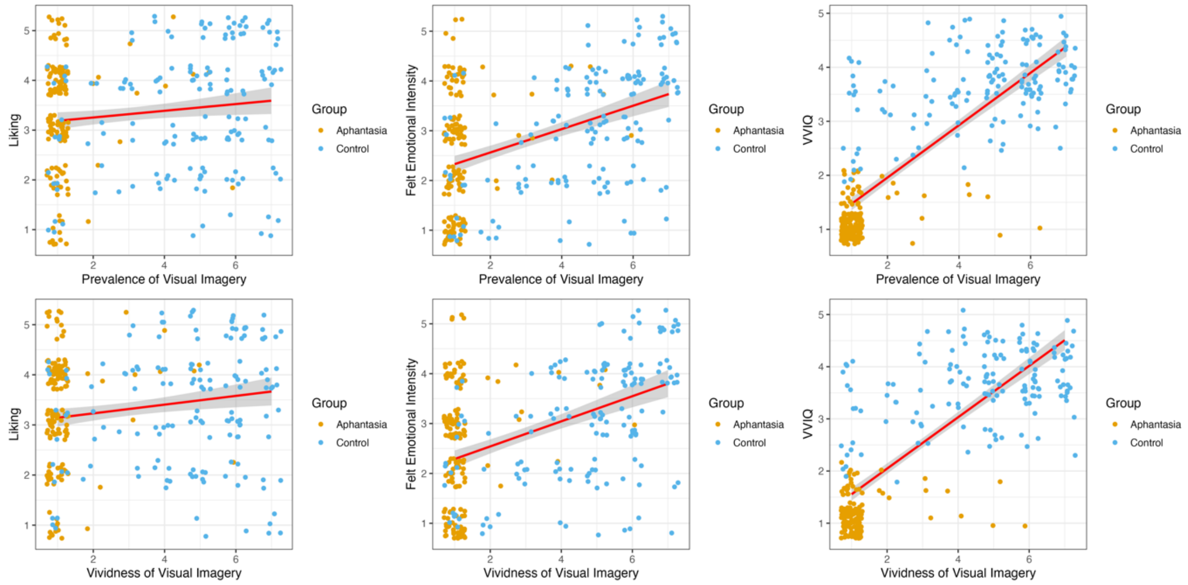
6.6.3. Links between behavioural measures

The model predicting prevalence of visual imagery revealed a significant positive association with felt emotional intensity (ß = 0.46, SE = 0.06, $t = 7.96$, $p < 0.001$, partial η2 = .37, 95% CI [0.25, 1.00]) and with VVIQ mean scores (ß = 1.10, SE = 0.15, $t = 7.40$, $p = 0.005$, partial η2 = .95, 95% CI [0.62, 1.00]), but no association with music liking (ß = 0.06, SE = 0.05, $t = 1.19$, $p = 0.253$).

Similarly, the model predicting vividness of visual imagery showed there to be significant positive relationships with felt emotional intensity (ß = 0.39, SE = 0.06, $t = 7.01$, $p < .001$, partial η2 = .16, 95% CI [0.10, 1.00]) and VVIQ (ß = 1.10, SE = 0.15, $t = 7.40$, $p = 0.015$, partial η2 = .96, 95% CI [0.52, 1.00]), but not with liking (ß = 0.09, SE = 0.05, $t = 1.96$, $p = 0.052$). Assessment of model assumptions using graphical means (histograms and Q-Q plots) indicated that residuals and random effects for both models satisfy normality. See Figure 6-2 for a visualisation of these results.

Figure 6-2

Regression lines with shaded 95% confidence intervals reflecting the relationships between music liking, felt emotional intensity, and VVIQ mean scores and the prevalence (top row) and vividness (bottom row) of visual imagery ratings. Data points represent ratings in response to the individual music excerpts from the aphantasic and control groups.



6.7. DISCUSSION: Study 5a

The aim of Study 5a was to provide a general impression of aphantasic listeners' affective and aesthetic experiences of music, in comparison with a subset of the general population group that I collected in Study 1 (Chapter 2). Hypothesis 1 stated that I would observe significant differences in visual imagery prevalence and vividness levels between aphantasics and the control group. This finding was supported and corroborates that visual imagery during music listening is highly prevalent amongst the general population (Study 1 in Chapter 2; Küssner & Eerola, 2019; Vuoskoski & Eerola, 2015).

The findings showed partial support for Hypothesis 2, which predicted that I would find significantly lower liking and felt emotional intensity ratings in aphantasic listeners than in the control group. No differences between groups in terms of music liking were found but I was able to identify differences in terms of felt emotional intensity levels, in line with past results (Dawes et al., 2020; Wicken et al., 2021). It is important to note though that despite showing differences, the result also demonstrated a very small effect size. Thus, this result should not be overinterpreted and may be specific to the current sample. That I was unable to find differences in terms of liking speaks to the idea that visual imagery is not necessarily a key determinant in the enjoyment of music but could simply be one general aiding factor (Belfi, 2019).

I was partially able to confirm Hypothesis 3, which stated that I would find a positive relationship between visual imagery and the aesthetic appeal of, and the felt emotional intensity towards, music in aphantasics and in the control sample. Overall, this was shown by the significant positive relationships between imagery prevalence and felt emotional intensity, even though there was no relationship with liking. Confirming Hypothesis 4, visual imagery prevalence and vividness were significantly related to VVIQ mean scores (Hashim et al., 2020; Küssner & Eerola, 2019).

In an additional set of analyses looking specifically at differences between the pure and minimal aphantasics and controls, I was able to confirm that the differences found between all aphantasics and controls still held when assessing differences between controls and these subgroups separately. The findings suggest that there may be little difference between aphantasics who report experiencing no visual imagery at all versus those who can experience very minimal levels of visual imagery in their imagined and affective responses to film music stimuli.

6.8. STUDY 5b: an exploratory assessment of aphantasics' emotional, everyday, and rewarding interactions with music

The purpose of Study 5b was to delve deeper into an exploratory assessment of how attenuated visual imagery abilities impact emotional engagement with music, as well as other key affective responses and functional uses of music more generally. If visual imagery is indeed important in the simulation of different emotional states, then one might expect this to affect an individual's ability to discriminate between different types of emotions that music can convey. Indeed, visual imagery has been thought to act as an "emotional amplifier" (Holmes et al., 2008). Given these ideas, I explored whether reduced visual imagery ability in aphantasics would also be related to a reduced musical emotion discrimination score (Aim 1).

Intuitively, it is possible that mechanisms other than visual imagery take precedence in engagement with music in the absence of visual imagery ability. Visual imagery is one of eight proposed mechanisms associated with music-induced emotions, others including brain stem reflex, rhythmic entrainment, evaluative conditioning, emotional contagion, episodic memory, musical expectancy, and aesthetic judgement. Thus, I explored the predominance of the remaining mechanisms of the BRECVEMA framework by including an adaptation of the MecScale (Juslin et al., 2014), a scale assessing listeners' experience of each emotion mechanism. As it is broadly unknown which particular mechanisms may facilitate emotional response to music in aphantasics, I aimed to explore the extent to which aphantasics exhibit interactions with music consistent with these mechanisms when compared to a control sample (Aim 2).

If aphantasics are less inclined to use visual imagination as a form of introspecting when listening to music, the uses that aphantasics have for music and how this contrasts with the general population would be a question of interest. A further aim of Study 5b was thus to examine the different functions that music holds for listeners by including the Adaptive Functions of Music Listening Scale (AFML; Groarke & Hogan, 2018) (Aim 3). The AFML includes factors relating to specific emotion-related everyday experiences, such as stress, anxiety, as well as general references to using music as cognitive and functional regulation related to reminiscence and sleep. I do not make specific predictions regarding these experiences as it is not clear how these music-related everyday experiences may be impacted by aphantasia compared to controls.

In Study 5b, I additionally examined patterns of reward from music using the Barcelona Music Reward Questionnaire (BMRQ; Mas-Herrero et al., 2013). As the findings in Study 5a demonstrated that aphantasics and controls showed no differences in music liking yet clear differences in terms of felt emotional intensity to music, I explored whether other rewarding aspects of listening to music would be affected by reduced visual imagery ability (Aim 4). Finally, I explored whether different dimensions of musical sophistication might be affected by aphantasia when compared to a control group, measured using the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) (Aim 5). It is possible that if differences are found then this might be reflected by weak effects, as it has previously been shown that there is only minimal influence of musical expertise on visual imagery content following an induction task (Herff et al., 2021), and that other types of imagery are generally more linked with high musical training than visual imagery (Talamini et al., 2022).

A final exploratory aim of Study 5b was, as for Study 5a, to evaluate potential differences between pure and minimal aphantasics and controls separately. My findings in Study 5a suggest there to be no marked differences between subgroups compared to when assessing aphantasics' responses to the musical excerpts as a whole. However, it remains a possibility that nuanced differences may be found when considering general functional and everyday uses of music.

6.9. METHODS

6.9.1. Participants

In the aphantasic sample, 29 individuals were recruited from the same pool of participants invited in Study 5a (who were recruited via dedicated Facebook groups; *pure* = 19, *minimal* = 10), and included in analyses (aged 33–80 years, 23 females, 5 males, 1 other, M = 57.6, SD = 13.7). This time, only *pure* and *minimal* aphantasics from the sample recruited in Study 5a were directly invited via email to take part in this study. In terms of location, 48% (n = 14) of respondents were residents of the UK, 24% (n = 7) of the USA, 14% (n = 4) of Canada, 7% (n = 2) of Australia, 3% (n = 1) of Germany, and 3% (n = 1) of China.

I collected a sample of participants from the general population matched according to age and gender using Prolific (an online platform that facilitates participant recruitment; www.prolific.co). In this control sample, 29 individuals were recruited (aged 33–71 years, 24 females, 5 males, M = 52.0, SD = 11.3). I also targeted individuals from the UK and USA, as these were the two most predominant countries of residence in the aphantasic sample. Of these, 93% (n = 27) were residents of the UK, and 7% (n = 2) were residents of the USA. All individuals within the control sample possessed summed VVIQ scores of 30+, making them an appropriate non-aphantasic control group due to their sufficient general visual imagery abilities.

Independent samples t-tests confirmed that there were no differences between the two groups in terms of age, $t(54.1) = 1.7$, $p = 0.096$, musical training, $t(54.6) = 0.7$, $p = 0.473$, or general musical sophistication, $t(53.9) = 0.0$, $p = 0.972$.

6.9.2. Ethics Statement

This survey has received ethical approval from the Ethics Committee of the Department of Psychology at Goldsmiths, University of London. Participants provided online written consent for the collection, analysis, and potential publication of their data in a scientific journal. They also received monetary compensation for their time.

6.9.3. Materials and stimuli

The online survey comprised a series of six questionnaires:

1) My adaptation of the MecScale (Juslin et al., 2014) assessed listeners' experiences of music with regard to eight BRECVEMA mechanisms, including rhythmic entrainment which was added to the framework later on (Juslin, 2013; Juslin & Västfjäll, 2008) as well as cognitive appraisal which was included by Juslin et al. (2014) in their original adaptation of the scale. Items in the original scale were phrased to probe responses to individual listening trials; however, each question was rephrased here to inquire into music listening more generally (see Table D.2 in Appendix D for the original and adapted items). As such, participants were provided with the following instruction to clarify that their ratings should be based on all past musical experiences: '*Please indicate the extent to which each of the following statements best reflects your past experiences with music listening*'. The scale was also converted from a dichotomous scale to a continuous 7-point Likert scale, from 1 (Strongly Disagree) to 7 (Strongly Agree), to allow for clearer comparison between the questionnaires used here.

2) The Adaptive Function of Music Listening Scale (AFML; Groarke & Hogan, 2018) is a 46-item psychometric tool for measuring functions to music based on 11 broad factors: Stress Regulation, Anxiety Regulation, Anger Regulation, Loneliness Regulation, Rumination, Reminiscence, Strong Emotional Experiences, Awe and Appreciation, Cognitive Regulation, Identity, and Sleep (e.g., 'I can escape from stressful situations by listening to music', 'Listening to music in bed helps me fall asleep', etc.). This scale was presented in a randomised order for each participant. This was assessed on a 5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). This scale possessed a Cronbach's alpha reliability score of $\boldsymbol{\alpha} = 0.94$.
3) The Barcelona Music Reward Questionnaire (BMRQ; Mas-Herrero et al., 2013) was included to gauge perceptions of musical reward on a 5-point Likert scale from 1 (Completely Disagree) to 5 (Completely Agree). It is a 20-item questionnaire that measures musical reward along five factors: Emotional Evocation, Sensory-Motor, Mood Regulation, Musical Seeking, and Social Reward (e.g., 'When I share music with someone I feel a special connection with that person', 'Music calms and relaxes me', etc.). This scale possessed a Cronbach's alpha reliability score of $\boldsymbol{\alpha} = 0.92$.
4) The full 39-item Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) was administered to record individuals' general musical sophistication along the 5 dimensions: Active Engagement, Perceptual Abilities, Musical Training, Singing Abilities, and Emotions (e.g., 'I sometimes choose music that can trigger shivers down my spine', 'I can sing or play music from memory', etc.). These were assessed on a 7-point Likert scale from 1 (Strongly Disagree) to 7 (Strongly Agree). This scale possessed a Cronbach's alpha reliability score of $\boldsymbol{\alpha} = 0.96$.
5) The Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973).

6) I administered the adaptive version (MacGregor et al., 2023) of the Musical Emotion Discrimination Test (aMEDT; MacGregor & Müllensiefen, 2019), which assesses one's ability to discriminate emotions in music, using the R package *psychTestR* (version 2.23.2). Participants are presented with two performances of the same melody (selected from a large bank of excerpts) by the same performer, each attempting to communicate a different type of basic emotion, and are asked to select the performance they think is being conveyed according to a certain target emotion in a two-alternative forced choice (2AFC) format (for example, participants would be presented with the word 'Happy' and then played two melodies consecutively, after which they will click on the rendition that they think conveyed that target emotion). Scores are calculated by the aMEDT program by summing the correct responses and converting the value to a percentage. This test allows for experimenter-chosen trial parameters, thus I opted to include 25 trials based on the test authors' findings of high correlations between true and simulated estimated test abilities using 25 trials ($r = 0.82$), and since correlations appeared to stabilise at 20 + test items. Using standard error of the mean (SEM) as a measure of reliability, the scale possessed an SEM of 0.3 when presenting 25 items. Guidelines for implementing this as a remote test were obtained from https://github.com/klausfrieler/EDT.

6.9.4. Procedure

Study 5b was built and presented using Qualtrics. Participants in both samples were first provided with explanations of the study purpose, instructions, and informed consent information. They were then asked to provide demographic details, such as an anonymous ID, age, gender, and country of residence. Participants then completed the six questionnaires described above in the following order: MecScale, AFML, BMRQ, Gold-MSI, VVIQ, and aMEDT2. The aMEDT2 included listening trials, and so participants were advised to ensure that they used good-quality headphones and had minimal outside disturbances (though it was not formally checked whether this advice was adhered to). The survey also required participants to visit an external site to carry out the aMEDT2 before returning to officially complete the study, thus the completion button on this instruction page was timed to only appear after approximately 2 mins to avoid careless clicking. To mitigate the possibility that participants would not return to the survey page to formally end the survey, the aMEDT2 was presented last in the questionnaire running order.

This online survey took approximately 30 mins to complete, and participants received £5 as compensation for their time.

6.10. ANALYSIS

Scores were computed for each questionnaire. aMEDT scores were missing from three participants due to technical issues. Statistical analyses were run in R (Version 4.2.2). Initial diagnostics of the data revealed non-normal distributions in at least two or more factors for each questionnaire, which transformations were not able to correct. To mitigate this, non-parametric tests or tests known for being robust against such issues were used where possible.

To test differences in emotion discrimination abilities, a non-parametric Mann–Whitney U test was run on aMEDT scores between the aphantasic and control groups. Multivariate linear models, which are robust against non-normality in dependent variables, were run to test potential differences between the aphantasic and control groups for their scores across the different questionnaires (MecScale, AFML, BMRQ, and Gold-MSI). In all models, factors from the questionnaires were entered as dependent variables while Group (aphantasic and control) was entered as the fixed effect. Pillai's Trace was used as the test statistic in the analysis involving the MecScale due to violations in homogeneity of covariance (Box's M = 69.32, $p < 0.001$).

These analyses were rerun to consider differences between the control group and the pure as well as minimal aphantasic groups. Independent samples t-tests maintained that there were no differences in age between the pure aphantasics (M = 57.11, SD = 14.21) and controls (M = 52.03, SD = 11.34), $t(32.52) = 1.31$, $p = 0.201$, and between the minimal aphantasics (M = 58.60, SD = 13.31) and controls (M = 52.03, SD = 11.34), $t(13.80) = 1.39$, $p = 0.185$.

6.11. RESULTS

6.11.1. Differences in emotional discrimination

A Mann–Whitney U test run on aMEDT scores showed that there was no difference between the aphantasic (M = 1.27, SD = 0.79) and control (M = 1.39, SD = 0.89) groups in terms of their emotion discrimination abilities of short melodies, W = 287, $p = 0.265$.

There were also no differences found between the pure aphantasic (M = 1.21, SD = 0.81) and control groups (M = 1.39, SD = 0.89), W = 184, $p = 0.210$, as well as between the minimal aphantasic (M = 1.38, SD = 0.79) and control groups (M = 1.39, SD = 0.89), W = 103, $p = 0.751$.

6.11.2. BRECVEMA mechanisms

Figure 6-3 presents each factor of the MecScale for the different sample groups and displays which mechanisms were most highly rated in the aphantasic and control samples' music listening experience. Multivariate linear model results revealed a main effect of Group ($F(8, 44) = 25.60$, $p < 0.001$, Pillai = 0.82, partial $\eta 2 = 0.82$) whereby the aphantasic group tended to score lower on the BRECVEMA factors than the control group. Given this main effect, follow-up linear models were run on each of the dependent variables, with a Bonferroni adjusted alpha level of 0.006 (see Table 6-1 for full model results).

Figure 6-3

Averaged scores of the MecScale factors for the aphantasic and control groups. Error bars reflect ± standard error of the mean.

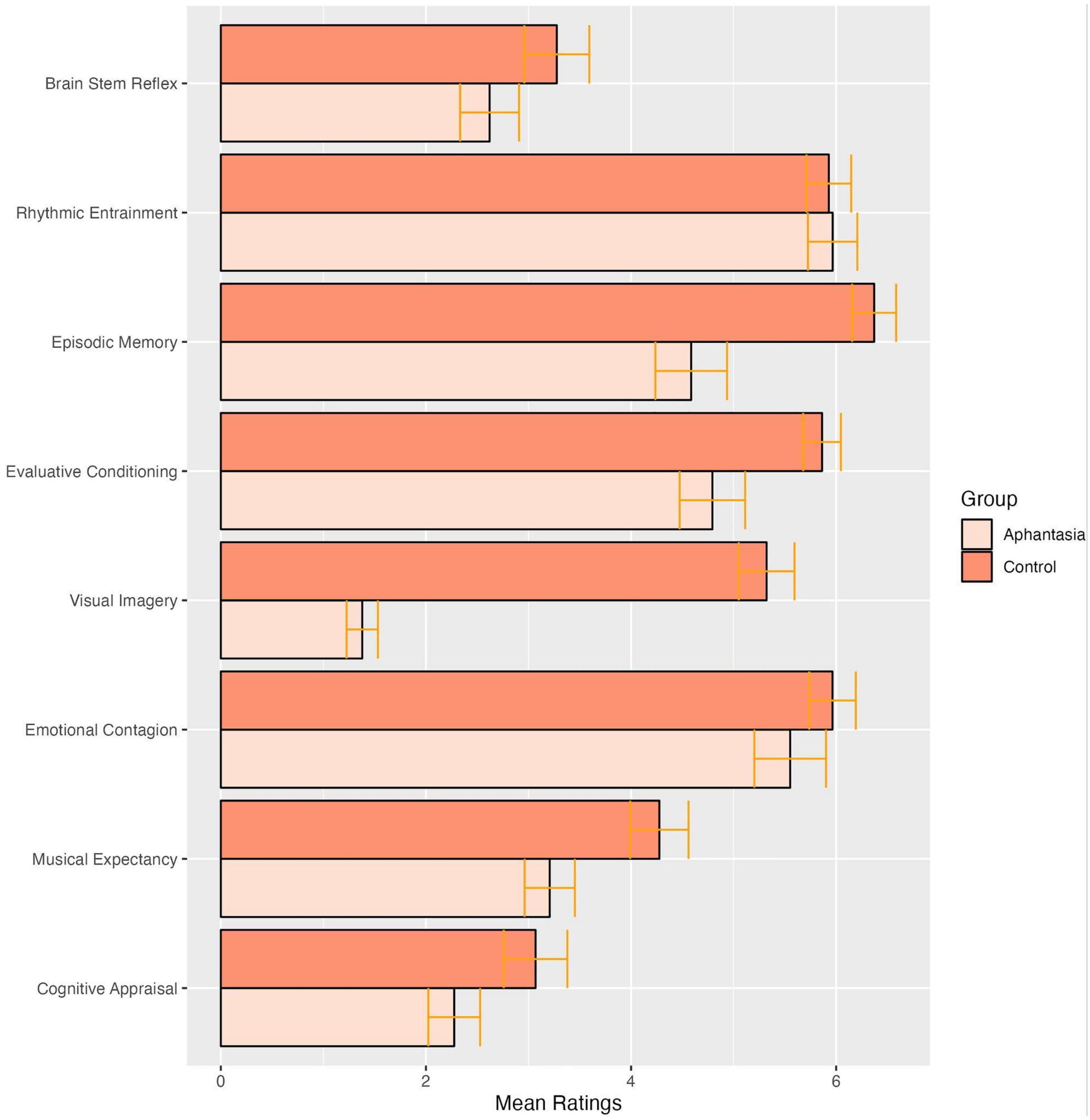


Table 6-1

Fixed effect estimates of multivariate linear mixed model on MecScale items between the aphantasia and control groups.

	ß	SE	95% CI	*t*	*p*	Partial eta^2
Brain Stem Reflex	0.75	0.46	[0.00, 1.00]	1.62	0.110	0.05
Rhythmic Entrainment	–0.09	0.34	[0.00, 1.00]	-0.26	0.794	0.00
Episodic Memory	1.71	0.44	[0.08, 1.00]	3.87	< 0.001*	0.23
Evaluative Conditioning	1.00	0.39	[0.01, 1.00]	2.52	0.015	0.11
Visual Imagery	4.08	0.30	[0.70, 1.00]	13.77	< 0.001*	0.79
Emotional Contagion	0.57	0.44	[0.00, 1.00]	1.31	0.195	0.03
Musical Expectancy	1.08	0.41	[0.02, 1.00]	2.66	0.011	0.12
Cognitive Appraisal	0.81	0.41	[0.00, 1.00]	1.96	0.055	0.07

Note. * < 0.006

Linear models revealed that the control group scored significantly higher than the aphantasic group in terms of Visual Imagery (ß = 4.08, SE = 0.30, $t = 13.77$, $p < 0.001$) and Episodic Memory (ß = 1.71, SE = 0.44, $t = 3.87$, $p < 0.001$). No other significant effects were found amongst the remaining mechanisms.

Further, multivariate linear model results between pure aphantasics and controls revealed a main effect of Group ($F(8, 34) = 25.75$, $p < 0.001$, Pillai = 0.86, partial $\eta 2 = 0.86$) whereby the pure aphantasic group tended to score lower on the BRECVEMA factors than the control group. Follow-up linear models revealed that the control group scored significantly higher than the pure aphantasic group in terms of Visual Imagery (ß = 4.35, SE = 0.32, $t = 13.72$, $p < 0.001$) and Episodic Memory (ß = 2.40, SE = 0.46, $t = 5.19$, $p < 0.001$), as well as in terms of Evaluative Conditioning (ß = 1.37, SE = 0.42, $t = 3.27$, $p = 0.002$) and Musical Expectancy (ß = 1.61, SE = 0.44, $t = 3.63$, $p < 0.001$). No other significant effects were found amongst the remaining mechanisms. See Figure D.3 in Appendix D for a visualisation of the group score differences and Table D.3 for full model results.

Finally, multivariate linear model results between minimal aphantasics and controls revealed a main effect of Group ($F(8,25) = 11.23$, $p < 0.001$, Pillai = 0.78, partial η2 = 0.78) whereby the minimal aphantasic group tended to score lower on the BRECVEMA factors than the control group. Follow-up linear models revealed that the control group scored significantly higher than the minimal aphantasic group in terms of Visual Imagery only (ß = 3.56, SE = 0.47, $t = 7.50$, $p < 0.001$). No other significant effects were found amongst the remaining mechanisms. See Figure D.4 for a visualisation of the group score differences and Table D.4 for full model results.

6.11.3. Everyday functions of music

Figure 6-4 presents average scores for each factor of the AFML for the two sample groups. Overall, individuals from the control sample rated music more highly than aphantasics in terms of its everyday uses. Multivariate linear model results revealed a main effect of Group ($F(11, 46) = 2.87$, $p = 0.006$, Wilks = 0.59, partial η2 = 0.41) whereby the aphantasic group tended to score lower on the AFML factors than the control group. Given this main effect, follow-up linear models were run on each of the dependent variables, with a Bonferroni adjusted alpha level of 0.005 (see Table 6-2 for full model results).

Figure 6-4

Averaged scores of the AFML factors for the aphantasic and control groups. Error bars reflect ± standard error of the mean.

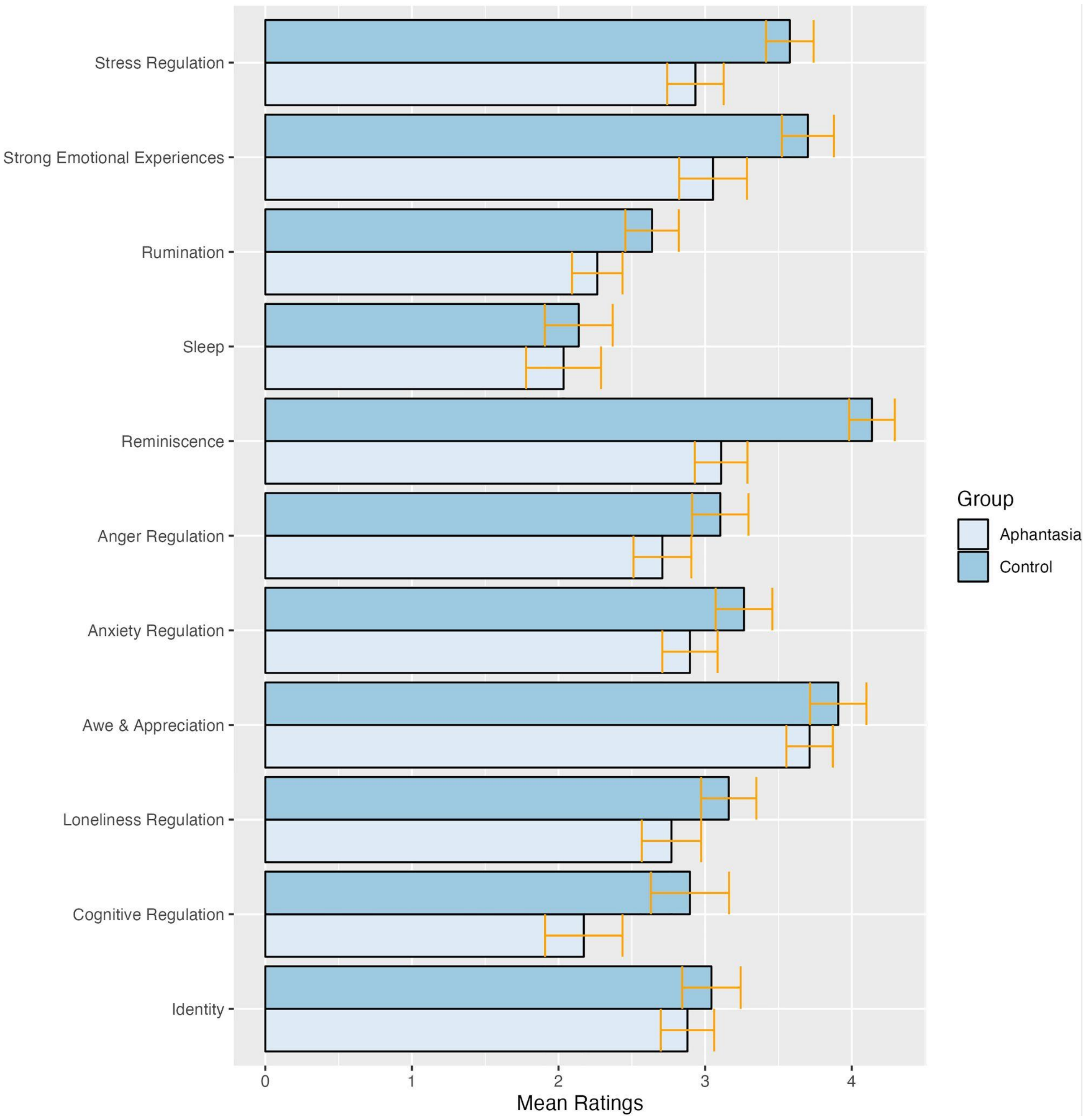


Table 6-2

Fixed effect estimates of multivariate linear mixed model on AFML items between the aphantasia and control groups.

	ß	**SE**	**95% CI**	***t***	***p***	**Partial eta^2**
Stress Regulation	0.64	0.25	[0.01, 1.00]	2.56	0.013	0.10
Strong Emotional Experiences	0.65	0.29	[0.00, 1.00]	2.22	0.030	0.08
Rumination	0.37	0.25	[0.00, 1.00]	1.49	0.142	0.04
Sleep	0.10	0.34	[0.00, 1.00]	0.30	0.765	0.00
Reminiscence	1.03	0.24	[0.10, 1.00]	4.33	< 0.001*	0.25
Anger Regulation	0.39	0.28	[0.00, 1.00]	1.43	0.157	0.04
Anxiety Regulation	0.37	0.27	[0.00, 1.00]	1.37	0.177	0.03
Awe & Appreciation	0.20	0.25	[0.00, 1.00]	0.79	0.436	0.01
Loneliness Regulation	0.39	0.28	[0.00, 1.00]	1.41	0.163	0.03
Cognitive Regulation	0.72	0.38	[0.00, 1.00]	1.93	0.059	0.06
Identity	0.16	0.27	[0.00, 1.00]	0.61	0.547	0.00

Note. * < 0.005

Linear models revealed that the control group scored significantly higher than the aphantasic group in terms of Reminiscence only (ß = 1.03, SE = 0.24, $t = 4.33$, $p < 0.001$). No other significant effects were found amongst the remaining factors after correcting for multiple comparisons.

Next, multivariate linear model results between pure aphantasics and controls revealed a main effect of Group ($F(11, 36) = 2.82$, $p = 0.009$, Wilks = 0.54, partial η2 = 0.46) whereby the pure aphantasic group tended to score lower on the AFML factors than the control group. Follow-up linear models revealed that the control group scored significantly higher than the pure aphantasic group in terms of Reminiscence (ß = 1.32, SE = 0.26, $t = 5.18$, $p < 0.001$) and in terms of Strong Emotional Experiences (ß = 1.07, SE = 0.30, $t = 3.58$, $p < 0.001$). No other significant effects were found amongst the remaining factors after correcting for multiple comparisons. See Figure D.5 in Appendix D for a visualisation of the group score differences and see Table D.5 for full model results.

Finally, multivariate linear model results between minimal aphantasics and controls revealed no main effect of Group ($F(11, 27) = 1.58$, $p = 0.161$, Wilks = 0.61, partial η2 = 0.39). Thus, follow-up models for the dependent variables were not run. See Figure D.6 for a visualisation of the group score differences.

6.11.4. Differences in musical reward

The multivariate linear model predicting the five factors of the BMRQ revealed no main effect of Group ($F(5, 52) = 1.53$, $p = 0.196$). Thus, follow-up models for each of the dependent variables were not run. See Figure D.7 in Appendix D for a visualisation of these factors.

Similarly, the multivariate linear model predicting the five factors of the BMRQ between the pure aphantasic and control groups revealed no main effect of Group (F(5,

42) = 0.70, $p = 0.623$). However, the model between the minimal aphantasic and control groups did reveal a main effect of Group ($F(5, 33) = 3.42$, $p = 0.013$, Wilks = 0.66, partial η2 = 0.34). Given this main effect, follow-up linear models were run on each of the dependent variables, with a Bonferroni adjusted alpha level of 0.01. The follow-up models did not reveal significant differences for any of the factors after correction. However, they reflected a tendency for the minimal aphantasic group to score lower than the control group in terms of the Musical Seeking factor (see Table D.9).

6.11.5. Differences in musical sophistication

The multivariate linear model predicting the five factors of the Gold-MSI revealed no main effect of Group ($F(5, 52) = 1.39$, $p = 0.243$). Thus, follow-up models for the dependent variables were not run. See Figure D.8 for a visualisation of these factors.

Similarly, the multivariate linear model predicting the five factors of the Gold-MSI between the pure aphantasic and control groups revealed no main effect of Group ($F(6, 41) = 1.92$, $p = 0.100$), as well between the minimal aphantasic and control groups ($F(6, 32) = 2.04$, $p = 0.089$). As a result, follow-up models for the dependent variables were not run.

6.12. DISCUSSION: Study 5b

One of the purposes of Study 5b was to understand the effects that a lack of visual imagery ability might have on emotional processing and engagement with music. In this case, I address Aim 1 by showing that aphantasia does not seem to affect emotion discrimination abilities. It is evident that while aphantasics may exhibit reduced emotional responses to imagined stimuli, this does not necessarily affect their ability to recognize and discriminate distinct emotional differences in perceived stimuli such as music.

There were, however, some notable group differences in terms of the BRECVEMA emotion induction mechanisms. Apart from visual imagery, aphantasics scored lower than controls in terms of episodic memory. The findings therefore add to Aim 2 in showing that mechanisms of music-induced emotions occurring in the visual domain or with the potential to be experienced visually (as in the case of memory) tend to be experienced less in aphantasics. Interestingly, these findings resonated with other significant differences between aphantasics and control listeners with regard to everyday music functions. In particular, the factor that was distinguished most strongly between groups was Reminiscence, the act of thinking back and remembering past events, contributing to Aim 3. Modulations in the experience of memory in aphantasics has been a focus of recent literature, with some studies highlighting that aphantasics exhibit deficient autobiographical memories (Dawes et al., 2020; Monzel et al., 2023; A. Zeman et al., 2020). The findings of Study 5b highlight that the general impairments in autobiographical memories reported by aphantasics further translates to their experience of music listening.

The differences in terms of memory-related factors found here are relevant, both in terms of the functions that music may serve and in terms of emotion-related mechanisms. Whilst aphantasics do not tend to exhibit any reductions in the process of consolidating working memory (Knight et al., 2022), my findings are in line with previous literature that has shown that visual imagery is strongly associated with the visual experience of memory (Handy et al., 2004; McKelvie & Demers, 1979; Pearson & Keogh, 2019). The ability to consolidate memories of events and then use it for later recall is a fundamental aspect of visual imagery; individuals with congenital aphantasia appear to show a reduced ability to relive their memories (Greenberg & Knowlton, 2014) and difficulties in simulating the future (Dawes et al., 2022).

Moreover, no differences between groups in terms of their musical reward along the BMRQ dimensions were observed (Aim 4), suggesting that a lack of visual imagery ability does not impede one's ability to derive rewards from music, including those pertaining to emotional rewards. No differences were further found between the aphantasic and control groups in terms of their musical sophistication index, as measured by the Gold-MSI (Aim 5).

Taking findings thus far, it appears that emotion discrimination and musical rewards are not majorly impacted by a lack of, or minimised, visual imagery abilities. However, aphantasics appear to generally report fewer instances of experiencing episodic memory and reminiscing past events in response to music.

The additional analyses examining pure and minimal aphantasics separately offer interesting results. Firstly, the finding of no differences between these subgroups and controls was maintained in terms of emotion discrimination abilities, musical reward, and musical sophistication. However, in addition to visual imagery and episodic memory, pure aphantasics appeared to exhibit markedly lower scores in evaluative conditioning (the enjoyment of musical stimuli paired with other positive or negative stimuli) and musical expectancy (emotions induced due to a confirmation or violation in expectations). Further, the findings suggest that, in addition to lower reminiscence in everyday functions of music, this subgroup further provided significantly lower ratings in terms of strong emotional experiences. Thus, I conclude that it is generally the more extreme levels of aphantasia that lead to more accentuated decreases in memory formation and emotional experiences during music listening. These results seem contradictory to the fact that none of the affect-related factors of the BMRQ and Gold-MSI were similarly negatively affected. Since the effect sizes are low, further replications would be necessary to explore the extent of the robustness and generalizability of these findings.

The fact that minimal aphantasics did not also exhibit significantly lower episodic memory, as in the pure and overall aphantasic groups, suggests that even low levels of visual imagery abilities may be enough to utilise music for the purpose of visually forming and recalling past events. This was further in line with finding no apparent differences in terms of the everyday uses of music in minimal aphantasics compared to controls.

6.13. GENERAL DISCUSSION

Across the research presented in this chapter, I compared aphantasics and controls with regard to differences in their experiences of music-evoked visual imagery, aesthetic appeal, and emotional response (Study 5a). I also explored patterns of experience and engagement with music in everyday life, as well as emotion discrimination ability (Study 5b). Specifically, the purpose of this research was to shed light on how emotional processes and responses to music are related to the experience of visual imagery. In Study 5a, I predicted that aphantasics and controls would demonstrate distinctly different visual imagery abilities (Hypothesis 1), as well as differences in their felt emotional intensity and liking towards music (Hypothesis 2). Relatedly, I predicted that, following evidence of group-level differences in the behavioural data, correlation analyses would reveal positive associations between visual imagery abilities and felt emotional intensity and liking in the aphantasic and control samples (Hypothesis 3). In terms of trait features, imagery prevalence was expected to be positively related to VVIQ mean scores (Hypothesis 4).

It was shown that the aphantasic and control groups did indeed show significant differences in their visual imagery abilities and did display differences in their ratings of felt emotional intensity towards the musical excerpts in Study 5a, albeit not in how much they liked them. These findings were further replicated when comparing pure and minimal aphantasics against controls separately. While my study tested this in the context of music listening, this finding is in line with previous results showing that a lack of visual imagery is associated with reduced feelings of fear when imagining negative situations (Wicken et al., 2021) as well as fewer experiences of intrusive memories (Dawes et al., 2020).

The results from Study 5a help to strengthen visual imagery's place in the BRECVEMA framework of music-induced emotions (Juslin, 2013; Juslin & Västfjäll, 2008). For example, it has been shown that emotive descriptions enhance the formation of visual imagery that heightens emotion induction during music listening (Vuoskoski & Eerola, 2015), while another study revealed that the active suppression of music-evoked visual imagery led to minor attenuation of emotional response to music (Hashim et al., 2020). The results are not capable of contributing to the discourse regarding the directionality of the causal relationship between visual imagery and music-induced emotion (Day & Thompson, 2019; Küssner et al., 2023; Vroegh, 2018), but I propose that the significant reductions in felt emotional intensity ratings in aphantasics could be the result of a lack of visual imagery ability. One interpretation for the reduced music-related emotional intensity is that, with impeded visual imagery abilities, aphantasics may have less ability to engage narratively with music (Day & Thompson, 2019; Küssner & Eerola, 2019; Margulis, 2017; Taruffi et al., 2017; see also Study 1 in Chapter 2). It is possible that this hinders one potential method for engaging emotionally with music; namely through the simulation of emotional visual scenes associated with the music. This poses further questions regarding aphantasics' reasons for listening to music (more specifically, film music, as was used here), and further research may benefit from adopting a nuanced and perhaps qualitative approach to addressing this question.

Following on from the finding that aphantasic and control individuals showed different emotional responses to musical excerpts, in Study 5b, I administered a series of questionnaires to explore whether aphantasia would be related to reduced emotion discrimination abilities (Aim 1) as well as to explore whether aphantasics exhibit any differences with regard to emotion induction mechanisms (Aim 2), functions (Aim 3), and rewards (Aim 4) of listening to music. I further explored group differences in terms of musical sophistication (Aim 5), as well as, once again, assessing potential differences when analysing pure and minimal aphantasics separately against controls. The purpose of this study was essentially to more broadly understand the various experiences that a lack of visual imagery may involve.

Regarding Aim 1, I found no differences between aphantasic and control individuals with respect to their musical emotion discrimination scores, despite finding an indication of lower felt emotional intensity in aphantasics in Study 5a. Similarly, no differences were found when analysing pure and minimal aphantasics separately. This new finding suggests that although past reports have shown dampened imagination-related emotional experiences in those with aphantasia (Wicken et al., 2021) and proposed the idea of visual imagery as an emotional amplifier (Holmes et al., 2008), aphantasia does not negatively affect the perception and discrimination of emotions expressed by music. However, some notable differences are the fact that the aforementioned studies mainly focused on the induction (not discrimination) of negative (i.e., fearful) emotions. It is plausible that visual imagery is not as necessary for the discrimination of perceived emotions expressed by music as for felt emotions induced by music. Further, with regard to feeling emotions, even in the context of a discrimination task, it is possible that positive emotions, like happiness, are not as impacted by reduced visual imagery abilities as negative emotions, like fear, are (Taruffi et al., 2017). This would be in line with neuroscientific evidence showing activation in the visual cortex when listening to fear-inducing (vs. happiness-inducing) music (Koelsch et al., 2013).

Differences in the MecScale factors were evident in my results. Unsurprisingly, visual imagery was the most distinctly affected mechanism of the MecScale. The second most affected mechanism was episodic memory, suggesting that aphantasia affects the ability to recall past events, a finding in line with a recent investigation by Dawes et al. (2022) who reported that aphantasia negatively affected the ability to re-experience the past and simulate future events when asked to remember past life events and imagine hypothetical future events. The results show that this can apply to memories previously formed in response to musical excerpts. The findings further suggest that this reduction in episodic memory is most pronounced in those with pure aphantasia, i.e., those who report experiencing no visual imagery at all in their day-to-day lives.

Interestingly, aphantasic, as well as pure aphantasic, listeners also reported significantly lower levels of reminiscence (AFML factor) in addition to fewer episodic memories (MecScale factor), when listening to music. Both findings are in line with the previously reported link between imagery and memory processes. The findings in Study 5b add emphasis to past assertions that visual imagery is important in the formation of memories (Otenen & Aydin, 2023). It is important to note that my results are not able to comment on aphantasics' abilities to form and recall episodic memories, but simply that they tend to report having fewer of these experiences while listening to music. Indeed, it has been proposed that aphantasics perform just as well as non-aphantasics when recalling the details of an episodic event (Arcangeli, 2023; Keogh, Wicken, et al., 2021; Zeman et al., 2015). Thus, further investigations should aim to implement a task that specifically targets the formation of episodic memory in response to music. However, it is interesting to note that the marked differences in memory-related processes with regard to the MecScale and AFML were not present when assessing minimal aphantasics, which may provide tentative support for the idea that the general ability to form visual imagery, even to a small extent, may lead to higher instances of memory formation and consolidation during music listening.

The analysis into pure aphantasics further illustrated that there were attenuations in terms of evaluative conditioning and musical expectancy (MecScale factors), as well as in terms of strong emotional experiences (AFML factor), extending results found in Study 5a regarding significantly reduced felt emotional intensity in aphantasics compared to controls. Regarding conditioning, it seems that aphantasic listeners tend to form fewer general positive and negative associations between music and other stimuli. Further investigation with regard to expectancy is also required, although it is plausible that reductions in imagery ability are related to difficulties in predicting incoming events during music listening.

6.13.1. Limitations and future directions

There are a few key limitations to note. While participants in Study 5a were instructed to provide felt emotional intensity ratings in response to the musical excerpts, it also remains a possibility that the emotional responses that participants provided were in response to their emotional reactions to memories of the movies, rather than of the musical excerpts themselves. Thus, future studies into aphantasics' relationship with music should aim to present music representing a musical genre with fewer connotations attached.

A further limitation of the research in this chapter concerns the use of the MecScale in Study 5b, a questionnaire used to assess responses corresponding to the BRECVEMA framework of music-induced emotions (Juslin, 2013; Juslin & Västfjäll, 2008). These mechanisms have received significant attention over the last decade, especially with regard to their true presence in our experience of emotions. While some studies have been able to show evidence for the influence of a handful of the mechanisms (namely brain stem reflex, emotional contagion, episodic memory, and musical expectancy; see Juslin et al., 2014, 2015), the current test could represent an over-simplification of the experiences of the different framework mechanisms.

The results suggest that differences observed between the whole aphantasic sample and controls tended to hold even when analysing the pure and minimal aphantasics separately. Nevertheless, I propose that findings about these particular subgroups should be interpreted with caution given their very small sample sizes.

Importantly, one should also remember that compared to Study 5a, Study 5b relied heavily on retrospection, a difficult task for many. It would be informative and fruitful to consider the implementation of an experience sampling methodology, where responses to and functions of music are reported in the flow of everyday life. Using this, one would be able to ascertain the situations in which aphantasics differ from controls with regard to memory and broad emotional experiences in the context of music listening.

A final limitation to consider is the extensive use of self-report measures. Visual imagery is widely recognized to be an elusive process to measure. Thus, attempting to capture an absence of it poses even further methodological constraints. Nevertheless, past studies incorporating implicit techniques to validate the experiences of aphantasia have been successful in revealing a lack of sensorial experiences in response to tasks that require visual imagery to be carried out (Kay et al., 2022). Such investigations aiming to obtain a more objective account of aphantasia are generally still in their infancy, but beneficial to strive for.

There are many avenues for future research. It would be important to employ listening tasks with which to address questions regarding music-induced emotions. These would be especially useful in distinguishing aphantasic responses while listening to musical stimuli differing in emotional content (Taruffi et al., 2017; Wicken et al., 2021). Additionally, utilising objective indices such as pupil dilation response (Kay et al., 2022) would be a valuable way to measure the sensory strength of music-evoked visual imagery. Finally, the current research only provided a general impression of the reasons and functions that music listening can hold in everyday contexts. It may be useful, as previously stated, to employ techniques such as the Experience Sampling Method (Juslin et al., 2008; Juslin & Laukka, 2004) for sending daily regular probes, given the richness and insight that can be gained from obtaining regular reports of music listening experience in day-to-day settings. Such developments would certainly gain significant headway in understanding the role of visual imagery in music, as well as discerning the unique experiences of listeners with aphantasia.

To what extent are the differences found in our emotion-related measures specific to the musical stimuli or simply reflect a general pattern of reduced emotionality? Given the past reported findings of reduced emotional experiences in response to imagined stimuli (Wicken et al., 2021), the results seem to corroborate this general trend in a different type of setting, i.e., music listening (Arcangeli, 2023; Dawes et al., 2020). This is not to say that aphantasics cannot derive enjoyment and experience a range of emotions in response to music that do not require imagination (Koelsch, 2014). It is possible that the lack of uniformity across measures regarding emotion-specific reductions in aphantasics compared to controls in Study 5b is due to the nature of the questions being asked. While Study 5a specifically instructed participants to form visual imagery to the music (supporting dampened emotionality to visual imagination), Study 5b measured multiple functional and everyday experiences to music more broadly (thus encouraging a cross-modal retrospection of past musical experience). Therefore, understanding the source of such differences in emotion warrants obtaining data on aphantasics' emotional responses towards non-musical as well as musical stimuli and situations.

6.13.2. Conclusion

The studies presented in this chapter add emphasis to the strong relationship that visual imagery and emotion have, and sheds light on some of the ways that listeners with aphantasia interact with music. It also confirms well documented links between imagery ability and memory processes. Past findings suggest that aphantasics tend to exhibit diminished emotional response to imagined stimuli. The findings from Studies 5a and 5b provide further support for this notion and show how limited imagery ability may impact the ways in which music is used and experienced.

CHAPTER 7. GENERAL DISCUSSION AND CONCLUDING REMARKS

In this chapter, the main findings of the five experimental works presented across Chapters 2 – 6 are summarised. General as well as experiment-specific limitations are considered. The implications of the empirical works are then outlined with regard to their relevance to the wider literature of imagination, visual imagery, and music listening. Finally, potential future directions of the research are suggested. In all, the present thesis offers novel insights into the contents, objective indices, and day-to-day applicability of music-evoked visual mental imagery.

7.1. SUMMARY OF KEY FINDINGS

The resurgence in research interest into visual mental imagery over the last decades may be due its prevalence in our everyday lives. Visual imagery permeates much of our experiences of listening to music, whether that be through the formation of autobiographical memories or through the facilitation of positive (or negative) emotions, making music an effective stimulus for inducing and measuring visual imagery under experimental settings.

The central aim of this thesis was to amalgamate key outstanding questions within music-evoked visual imagery research that would improve our understanding of the phenomenon. Through the combination of behavioural and neurophysiological techniques, the five empirical investigations described throughout this thesis help to provide a fuller understanding of the modulatory impact of music on visual imagery.

7.1.1. Music-evoked visual imagery is predominantly story-like and relatively consistent.

Study 1 in Chapter 2 adopted an in-depth qualitative approach to extend what is known about music-evoked visual imagery. It aimed to build upon previous papers that have solely addressed the contents of what we imagine to music (Dahl et al., 2022; Küssner & Eerola, 2019) by explicitly addressing the question of whether visual imagery is necessarily idiosyncratic or whether music possesses semantic affordances that promote shared imagined content across listeners (Margulis, Wong, et al., 2022). Specifically, the main aims of Study 1 were (1) to formulate a hierarchical thematic framework of the most common visual imagery content formed to music, and (2) to understand how consistent a listeners' visual imagery was when compared to their own experience across time points and across the experiences of other listeners.

Participants were asked to provide unrestricted descriptions of any visual imagery they may have experienced in response to the musical excerpts, giving them the flexibility to refer to visual as well as non-visual imagery in their desired level of detail. The thematic analysis revealed that *Storytelling* was the most predominant form of music-evoked visual imagery, which contained subthemes pertaining to motor actions as well as settings and locations. Additional higher-order themes referred to *Associations*, references to memories, emotion, and other imagery, and *References*, details regarding other types of media. These results align closely to categories found by the aforementioned studies, as well as highlight that experiences to music are inherently multi-modal. They also strengthen visual imagery's role in the BRECVEMA framework given that emotion was a sizable subtheme of the *Associations* category. Furthermore, the consistency analysis showed that while participants were more consistent in themselves across time points, this level was marginally higher than the level of consistency shared across individuals. These findings extend the work of Margulis, Wong, et al. (2022) by providing direct comparisons of the themes referred to by listeners.

7.1.2. Different subtypes of music-evoked visual imagery possess distinguishable oscillatory signatures.

Despite the recent growth of research into music-evoked visual imagery, few studies have sought to tease apart the features that lend it its richness. It is evident through the observation of listeners' descriptions of their visual imagery that imagery can be standalone and static, or changeable, sequential, and moving. It would appear visual imagery may implicate motoric processes that lead to the latter type of visual imagery. However, this had not been explored previously.

The purpose of Study 2 in Chapter 3 was therefore to use EEG to examine the neural correlates underlying static and dynamic visual imagery formed in response to twenty-four 30-second excerpts of positive, neutral, and negative music. Participants listened to each musical excerpt and then rated the extent to which they experienced static imagery followed by dynamic imagery in response. Results revealed that both forms of visual imagery were associated with alpha power suppression in the parieto-occipital regions of the brain, in line with past research finding such a relationship and reinforcing visual imagery's *visual* nature (Cooper et al., 2003; Fachner et al., 2019; Sousa et al., 2017; Xie et al., 2020). However, dynamic imagery's further link with enhanced gamma power suggests that dynamic forms of music-evoked visual imagery are a complex phenomenon. Dynamic imagery's association with a desynchronisation and synchronisation with beta power is a further reflection of its recruitment of motor regions of the brain (Menicucci et al., 2020; Zabielska-Mendyk et al., 2018) and potential involvement of cross-modal processing (Villena-González et al., 2018).

In an additional sequence of EEG investigations, the current thesis also presents neural data that distinguishes between different intentionality states of music-evoked visual imagery. Since visual imagery, and, indeed, other forms of spontaneous thought in general (e.g., mind wandering, dreams), are postulated to differ in terms of their levels of deliberate constraint (Christoff et al., 2016), one question that arose concerned the conditions in which one can “induce” spontaneous and deliberate states of thought (Taruffi & Küssner, 2019). Studies 3 (Chapter 4) and 4 (Chapter 5) attempted to measure the predominance of spontaneous and deliberate forms of music-evoked visual imagery by utilising a probe-caught methodology during extended listening, which involved sending thought probes at quasi-random intervals throughout the course of 15–20-min listening trials. Study 3 sought to compare four listening conditions that differed in terms of familiarity (familiar vs. unfamiliar) and relaxation potential (relaxing vs. non-relaxing), whereas Study 4 focused on relaxation potential and compared three listening conditions comprising a relaxing Ambient track, a non-relaxing Techno track, and an active control podcast track. Throughout listening, participants were instructed to listen with eyes closed (suggested to enhance introspection; see Herff et al., 2022) and to open them when required to respond to probes asking if they had just experienced any visual imagery (responses: yes/no) and if so, how intentional they thought the experience was (responses: spontaneous/deliberate).

The results of Study 3 indicated spontaneous visual imagery to be the more prominent manifestation of music-evoked visual imagery, as apparent in self-report ratings. Behavioural ratings also showed visual imagery rates to be highest in response to music that was familiar and relaxing. In Study 3, clusters of oscillatory activity also suggested that occipital alpha suppression was more apparent in moments prior to spontaneous imagery probes than deliberate visual imagery probes. Similarly, self-report results of Study 4 also showed spontaneous visual imagery to be more highly experienced than deliberate visual imagery (especially in response to the relaxing Ambient track).

Across these three EEG experiments, it has been repeatedly replicated that forming music-evoked visual imagery leads to activation in the visual cortex, contributing to the current gap in neuroscientific investigations using music as an induction stimulus (Fachner et al., 2019). In all, Studies 2 – 4 (Chapters 3 – 5) demonstrate that, not only are participants conscious of and reliable in reporting on their own meta-cognitive states (Pearson et al., 2011; Seli et al., 2015), but these different visual imagery states are also measurable using objective neural indices.

7.1.3. Music-evoked visual imagery and different types of affective experiences are positively linked.

An additional key aim of Study 4 (Chapter 5) was to examine music-evoked visual imagery's potential role in alleviating feelings of stress as caused by a stress induction task. There is much evidence concerning the therapeutic benefits of forming visual imagery both within music-based therapies and as a strategy for improving emotion and self-regulation (Herff et al., 2022, 2023). Thus, in Study 4, participants took part in a 5-minute stress induction task followed by 15 mins of listening. Visual imagery was monitored using thought probes sent throughout the listening trials, and listeners' emotional states were measured before and after stress induction as well as after track listening using self-reports and SCR.

The results demonstrated that visual imagery evoked by music was effective in reducing feelings of stress, especially while listening to the relaxing Ambient track, evidenced by self-reported reductions in negative affect as well as reduced SCR. Further, music-evoked visual imagery was found to drive self-reported stress reduction, but not SCR reduction. My results show support for past assertions that mental imagery could act as a form of underlying mechanism in promoting wellbeing, potentially through forming positive projections of the future (Panteleeva et al., 2018). It also provides support for the utility of guided imagery and music interventions by reinforcing that visual imagery formed in response to pre-selected relaxing pieces of music is more likely to improve behavioural and physiological wellbeing (Fachner et al., 2019).

In Studies 5a and 5b (Chapter 6), the main aims were to explore the extent to which aphantasics, those with a reduced ability to form visual imagery, would exhibit weakened emotional response towards musical stimuli, as well as to explore the implications that reduced imagery ability would have on day-to-day interactions with music. These surveys demonstrated that while aphantasics reported weaker emotional intensity to music than controls, impaired visual imagery did not affect how much they liked the music. One possibility is that aphantasics reduced emotional response to the stimuli as a result of an inability to interact narratively with the music (Margulis, 2017). Nevertheless, the results reinforce music-evoked visual imagery's relationship with emotion induction (Day & Thompson, 2019; Hashim et al., 2020; Juslin, 2013, 2019; Vuoskoski & Eerola, 2015), and suggests that while vivid visual imagery may be linked to increased music liking (Belfi, 2019), it does not determine it. Despite my aims, an unexpected finding however was that listeners with aphantasia displayed marked reductions in levels of reminiscence and episodic memories formed in response to music, which poses new questions for the relationship between autobiographical memories and visual imagery.

7.2. LIMITATIONS OF THE RESEARCH

With specific limitations having already been mentioned at the ends of empirical Chapters 2 – 6, there remain only a few broader important limitations of the research to consider.

7.2.1. Music-evoked visual imagery and emotion induction

Findings presented throughout Studies 1 and 5a (Chapters 2 and 6, respectively) have, to varying capacities, re-established that visual imagery and emotion induction during music listening are related processes. Specifically, emotion-related content seems to integrate seamlessly into the content descriptions of music-evoked visual imagery. Also, aphantasics' reduced ability to generally visualise seems to be associated with reduced felt emotional intensity to the film music presented to them. Despite this evidence, however, the current thesis cannot contribute to the current growing discourse on the causal relationship between visual imagery and music-induced emotions (Day & Thompson, 2019; Vroegh, 2018, 2021), which currently possesses inconclusive directionality claims that seem to oppose the initial perspective of the BRECVEMA framework that proposed visual imagery to act as a mediating mechanism between music and emotion induction (Juslin, 2013; Juslin & Västfjäll, 2008).

7.2.2. Operationalisation of visual imagery

A general limitation in the current research into visual imagery during music listening, which has inadvertently impacted the current research, is that there are no definitive operationalised metrics for music-evoked visual imagery content. In Study 3 (Chapter 4), the content of visual imagery, while not analysed in the current thesis, was operationalised by employing continuous ratings based on the prominent themes that were present in Study 1 (Chapter 2). This may have presented a haphazard way of delving into the qualitative nature of visual imagery that lacked in depth, in my attempt to avoid more extensive analyses of content descriptions.

Furthermore, the current research alternated between the use of the VVIQ and the Psi-Q as an indicator of a participant's general visual imagery ability. However, the use of the Psi-Q consistently yielded inconclusive results compared to those provided by the VVIQ in Studies 1, 5a, and 5b. It is possible that each index may differ in their sensitivity to individual differences in imagery processing. Additionally, the VVIQ and Psi-Q may interact differently with musical stimuli, leading to variations in the perceived vividness and clarity of imagery experiences. For example, certain music characteristics, such as tempo, rhythm, and instrumentation, may elicit stronger visual imagery responses compared to auditory or tactile imagery. Nevertheless, while comprehensive in their assessment of various sensory modalities, the VVIQ and Psi-Q may not specifically address the unique ways in which music influences sensory imagery, demonstrating a need for a more domain-specific test of visual imagery (Küssner & Eerola, 2017).

7.2.3. Statistical techniques

The results from Studies 5a and 5b in Chapter 6 illustrated that aphantasics experience markedly lower felt emotional intensity to music compared to general listeners, and also differ in how they utilise music in everyday contexts. However, there are other analytical techniques that could be used throughout these studies in order to address the questions differently, ones that are equally, if not more, robust to sample biases. Indeed, the sample of aphantasics recruited here, expectedly, led to a majority of negatively skewed distributions in response to the majority of indices. I opted to use linear mixed effect models in addressing the relevant research questions due to its ability to tolerate non-normal distributions, and thus reveal significant differences between groups. However, given that aphantasia represents an absence in the sensory experience of visual imagery, an interesting question to address would be whether aphantasics report a significant lack of visual imagery. To this end, mixed models based on Bayesian inference would be an effective example of a robust statistical technique to examine this absence.

7.2.4. Range of musical stimuli during listening tasks

A noteworthy limitation to consider concerns the breadth of the stimuli presented both in Study 1 (Chapter 2) and in Study 5a (Chapter 6). While the purpose of Study 1 was to provide a detailed hierarchical framework of the most commonly experienced visual imagery content experienced in response to music, it cannot be considered systematic in its current state. Namely, the visual imagery framework was developed in response to only three musical excerpts and were taken from film soundtracks (Eerola & Vuoskoski, 2011). My choice of the film genre was partly to ensure that participants were free to provide rich accounts of their visual imagery in response to a programmatic selection of tracks, as well as giving myself considerable power with which to analyse consistency, especially when selecting such a low number of listening stimuli. Such decisions mean that the stimuli fall short of being considered entirely comprehensive. It is further a possibility that the predominance of storytelling and media references in the framework was an artefact of musical cues present in my chosen tracks that were associated with the development and changes found in film scenes.

However, these issues could be considered negligible. Margulis (2017) found that even when unprompted, individuals listening to instrumental classical music made references to film and television in their narrative open-text reports. Similarities between my study and theirs in revealing a high volume of media references may lie in the fact that my stimuli were predominantly orchestral. Nevertheless, future investigations should aim to incorporate music from genres with contrasting compositional characteristics and instrumentation. One piece of work by Markert and Küssner (2021) which utilised ambient music, typically incorporating non-instrumental synthesised compositional techniques, found that while participant reports presented a significant amount of storytelling visual imagery in response to both ambient and classical music excerpts, reports in response to the ambient music tracks also predominantly featured abstract visual imagery (i.e., non-specific images, such as geometric shapes and colours). These findings emphasise the variability in visual imagery that differences in music genre could lead to, and hints at the potential for the framework presented in Study 1 to be representative of most music genres with further refinement.

In the case of Study 5a, within the context of aphantasics, it should be once again emphasised that it is a novel finding that a reduction in visual imagery ability was in turn able to lead to reduced emotional intensity in response to the music. However, the conclusions that could be made with regard to the emotional responses of aphantasic listeners were limited by the use of excerpts conveying only three basic emotions. This could have hindered the range of pieces that listeners were able to reflect on; for example, including excerpts conveying sadness could enhance emotional effects, given the wide literature surrounding the paradoxes and pleasures of listening to sad music (Eerola et al., 2016; Taruffi & Koelsch, 2014; Vuoskoski et al., 2012).

7.3. BROADER IMPLICATIONS OF THE RESEARCH

The results of the current thesis provide a number of interesting contributions, which relate both to imagination research as a whole, and to the music domain more specifically. The current section will summarise a few important implications that arise from the empirical works presented throughout this thesis and will end with broader general implications related to imagination, visual imagery, and music listening.

7.3.1. Implications of the presented empirical works

The results of the experiments presented throughout this thesis pose key implications on the phenomenology of music-evoked visual imagery and offer novel insights into techniques that can be used to target music-evoked visual imagery in future research.

7.3.1.1. Relevance of using more innovative methodological approaches in visual imagery research

The present research has demonstrated the applicability and usefulness of a range of different methodologies to answer nuanced questions relating to music-evoked visual imagery. Firstly, in Study 1 (Chapter 2), the question of how consistent music-evoked visual imagery tends to be within and across listeners was addressed using the Jaccard coefficient index, a simple yet effective measure of overlap and an explicit approach for gauging the consistency of visual imagery content at code-level. The effectiveness of this builds upon other techniques used in recent research to address similar questions (i.e., cosine similarity index, Margulis, Wong, et al., 2022), and offers one optional approach for the investigation of text-based content similarity that is easy to implement and consumes minimal computational load.

Furthermore, as previously noted in Studies 3 and 4 (Chapters 4 and 5, respectively), the current research provides evidence for the applicability of the probe-caught methodology, whereby participants answer thought probes sent to them throughout an extended listening block, for the study of music-evoked visual imagery. This is the first instance in which this paradigm has been administered to investigate the occurrence and intentionality of music-evoked visual imagery and was valuable for providing real-time reports of visual imagery and minimising retrospective bias that has applied to most previous visual imagery research. With further refinement to the administration of the probes, there is potential to gain more contextual information by aligning the occurrences of music-evoked visual imagery to the specific acoustic events across the time-course of the music, enabling associations to be drawn between visual imagery content and influential musical features.

Finally, some of the research presented throughout this thesis has demonstrated the relevance of using an objective measurement technique such as EEG to assess the neural underpinnings of music-evoked visual imagery, the outcomes of which I will outline in more detail in the next section (*7.3.1.2*).

7.3.1.2. Clarifying the neural correlates of music-evoked visual imagery

The results derived from the EEG investigations presented across Studies 2 – 4 (Chapter 3 – 5, respectively), in addition to the one other music-related EEG study of visual imagery to date, contribute to establishing the role of alpha power suppression in the posterior regions of the brain as a robust index of visual imagery experience. It, however, extends it to implicate additional frequency bands and brain areas that dissociate between music-evoked visual imagery that varies in terms of content and intentionality. Specifically, Study 2 showed that, in addition to the posterior alpha suppression found during static and dynamic formation, dynamic imagery may further implicate motoric processes and increased cognitive load as reflected with enhanced beta and gamma activity. During extended listening in Study 3, results revealed spontaneous visual imagery to be associated with alpha and gamma power suppression. Across Study 4, deliberate imagery was associated with theta, alpha, and gamma power suppression, while visual imagery was in general linked to suppression in delta.

While the current research has offered a number of novel approaches with which to study the phenomenology of music-evoked visual imagery, with further work, research can continue to establish the efficacy of other uncharted objective measurements for exploring the association between additional aspects of music-evoked visual imagery formation and differing physiological and sensory experiences (e.g. using eye-tracking techniques, Kay et al., 2022).

7.3.1.3. Clarifying relationships between visual imagery and other cognitive processes

The present work demonstrates visual imagery's purported link with other cognitive experiences, namely: mind wandering and autobiographical memories. With regard to mind wandering, an important finding from Studies 2 – 4 (Chapters 3 – 5) is that of posterior alpha power suppression reliably accompanying music-evoked visual imagery (Fachner et al., 2019; Sousa et al., 2017; Xie et al., 2020); this, given that mind wandering is indexed by *enhancements* in alpha power (Polychroni et al., 2022). One potential reason for this fundamental distinction may be differences they tend to show in terms of alertness and task relatedness. A recent systematic review by Kam et al. (2022) on the electrophysiological markers of mind wandering associated alpha power increases with tasks related to internally-directed mind wandering and shifts away from the external tasks. In contrast, alpha suppression mainly coincided with externally-directed mind wandering during an internally-focused task (e.g., breath focus task; Braboszcz & Delorme, 2011). In this case, visual imagery could constitute an internal task where music is in turn the external task being attended to. In any case, differences between visual imagery and mind wandering are likely to be related to differences in executive control and alertness (Irving, 2016; Kane & McVay, 2012).

Furthermore, regarding autobiographical memories, the results from Study 5b (Chapter 6) reaffirm that visual imagery and memory are interlinked in fundamental ways. Visual imagery is evidenced to be a key modality in which memory is actualised in the mind and aids in encoding by providing vivid contextual and sensorial information that enhances the respective memory (Lenormand et al., 2024). Further, memories evoked in response to music are often reported to be experienced in the visual domain and are described in terms of a more embodied and social nature in comparison to memories in response to non-musical cues (Jakubowski et al., 2021; Jakubowski & Ghosh, 2021). Given this overlap, it is possible that episodic memory and visual imagery may not be completely separate entities (Jakubowski, 2022), posing implications for Juslin's (2013) framework for music-induced emotions that described visual imagery and episodic memory to be distinct mechanisms.

7.3.1.4. Corroborating the therapeutic potential of music-evoked visual imagery

Study 4 (Chapter 5) of this thesis has endorsed the notion that visual imagery can play a role in improving wellbeing and therapy, and specifically that it is a potentially useful strategy for stress and anxiety reduction. Visual imagery, conjured in response to music, was associated with increased relaxation after a stressful task, albeit mainly in terms of subjective stress ratings. It is possible that administering the probe-caught methodology for assessing this type of question led towards more cognitive freedom with which to form visual imagery at will, thus emulating the meditative environment of some music therapy sessions. However, in order to understand the extent of the beneficial effects of music-evoked visual imagery, the induction of negative emotional states should be considered.

A key conclusion from Studies 5a and 5b (Chapter 6) is that an absence of visual imagery is associated with reduced music-induced emotional intensity, a finding that is in line with other studies showing that disrupted visual imagery during music listening in turn leads to attenuated felt emotional response (Hashim et al., 2020). These results pose implications for therapy, as it means that certain types of interventions that leverage visual imagery experience to music for improving emotional and mental wellbeing, such as the Guided Imagery and Music (GIM) therapy intervention (Bonde, 2007; Bonny, 1986), are not appropriate for individuals with aphantasia. It is possible that other music-based psychotherapeutic interventions, including ones that harness other modes of non-visual imagery (for reviews, see Finch & Moscovitch, 2016 and Yinger & Gooding, 2015), could be more beneficial in the rehabilitation of individuals with reduced visual imagery abilities

7.3.2. Implications for imagination research

It has been argued that there is a strong interplay between visual imagery and imagination, but that they are, regardless, independent processes with their own purposes and capacity for intentionality. In this view, the imagination of an object or scene is usually rooted in inferences based on reality that make it accessible to others, that do not implicate fictional reinterpretations of an object or scene that others cannot share (Nishizaka, 2003). Visual imagery, on the other hand, possesses more contextual freedom, and constitutes a type of imagination that is based entirely on the intentions and prior experiences of the individual.

Nevertheless, one common thread in current literature is that imagination can be conceptualised as a holistic process that draws from all facets of sensory and perceptual cognition.

The results of the current thesis propose that visual imagery serves as a building block and form of expression and sensory manifestation of imagination. This is in line with Abraham's (2016) framework of the classifications of imagination. There, the aesthetic experiences derived from music listening, corresponding to interoceptive and emotional awareness, was crucial in influencing the content of the imagery (Abraham, 2016, 2020). While this thesis cannot argue for a distinction between what is visual imagery and what is imagination, there is overlap in that individuals are variably inclined to draw upon true instances from lived experiences or to fabricate static or on-going scenes in response to the presented music. Specifically, many of the phenomenological attributes that are found to arise from visual imagery in terms of content, its interaction with other sensory and emotional domains, and its neural mechanisms, may be similar for imagination. Against this context, the vividness and richness of visual imagery, guided by the intrinsic or associative features of musical stimuli, may contribute to the depth and detail of imagined scenarios.

7.3.3. Implications for visual imagery research

Building on this, it is evident that visual imagery is highly amenable to the evoking stimulus. By studying music-evoked visual imagery, researchers can contribute to a deeper understanding of how the mind processes and represents sensory information. One important implication of the current research is that the qualitative analysis of the contents of music-evoked visual imagery yielded rich insights into the cross-modality of mental imagery. Indeed, even when explicitly instructed to report on their visual imagery experience, participants' descriptions were heavily interspersed with details spanning other mental imagery domains as well as more abstract concepts. Content descriptions were varied in terms of their depth and vividness, but nevertheless, participants included a mixture of pre-conceptual mappings and metaphorical language with which to express and expand upon their visual imagery.

The findings lend support to the notions that music evokes a so-called metaphorical space and time (Di Bona, 2017) and a semantic dimension that affords narrative engagement (McAuley, Wong, Mamidipaka, et al., 2021), rendering imaginings of the musical events that are relatively shared across listeners within similar cultures (Margulis, Wong, et al., 2022; McAuley, Wong, Bellaiche, et al., 2021). It would be interesting for future work to investigate how the evocation of visual imagery generalises to other auditory stimuli. Investigating how different types of stimuli evoke visual imagery would provide valuable insights into the mechanisms underlying cross-modal processing, which would contribute to our fundamental knowledge of how the brain integrates information from multiple sensory modalities to form coherent perceptual experiences.

7.3.4. Implications for music listening research

One of the outcomes of this thesis was gaining further understanding on the propensity for music, an abstract and aesthetic stimulus with complex intrinsic qualities that render rich mental and emotional experiences, to modulate the features of visual imagery. Music’s widely acknowledged ability to induce a wide range of emotions, from joy and excitement to sadness and nostalgia, was evident throughout due to its prevalence in the imagery content provided by listeners.

Furthermore, the current thesis provides implicit support for music’s ability to induce altered states of consciousness, characterised by heightened awareness, altered perception of time and space, and profound emotional and spiritual experiences. One such example is that of absorption, an effortless form of attention involving loss of reality orientation and awareness of body and time (Vroegh, 2022), with visual imagery being one related experiential attribute (Pekala, 1991; Vroegh, 2021). Furthermore, with its varying degrees of complexity, music's ability to influence imaginative states allows one to "un-focus from the outer world and to be brought into the 'jungle gym' of cross-modal correspondences" (Fachner, 2022, p. 205). Visual imagery accompanying music can serve as a gateway to these transcendent states, allowing individuals to transcend their everyday reality and access deeper layers of consciousness. In all instances, one remaining goal from this area of research is understanding the factors that influence the various states of consciousness, such as the extent to which variability in the features of music influences mental imagery (as will be discussed in the following section).

7.4. FUTURE DIRECTIONS

7.4.1. Examining the influence of music's acoustic features on visual imagery formation

While the research presented throughout the thesis offered insights into the types of music – differing across levels of valence, genre, familiarity, and length – that can lead to visual imagery, it does not touch on the particular intrinsic features of the music that may be having a direct influence. Indeed, very little research has explicitly addressed this question, as only until recently have researchers hypothesised links between visual imagery content and certain cues within the music (Juslin, 2013; Juslin & Västfjäll, 2008; Vuoskoski & Eerola, 2015); for instance, that differences in musical contrast (as defined by systematic 'highs' and 'lows' in features such as dynamics, rhythm, and articulation) can influence the likelihood that an individual hears music in terms of a narrative (Margulis, 2017). In a footnote, Juslin (2019, p. 337) briefly explains that musical characteristics such as repetition, tempo, and predictability in melodic, harmonic, rhythmic elements could all play a role in stimulating vivid visual imagery. Despite a lack of systematic investigation – that spans across musical genres – on this topic, some studies have targeted particular acoustic features. For instance, Herff et al. (2021) found that tempo successfully mediated the relationship between the auditory stimuli and participants' imagined content, with faster tempo leading to less imagined time passed and distance travelled. Additionally, in a systematic analysis of how the emotionality and acoustic features of instrumental music influence mental imagery content, Groves et al. (2023) found that certain acoustic features such as pitch and loudness have a nuanced impact on the characteristics of imagined movie scenes. Further findings in this regard will surely not only shed light on the particular acoustic musical features that lead to certain types of visual imagery, but also help to make the hierarchical framework in Study 1 (Chapter 2) representative of a much richer pool of musical genres and compositional styles.

7.4.2. Understanding the temporal dynamics of music-evoked visual imagery

The research presented in Chapters 3 – 5 hinted at the idea that visual imagery can be highly variable over time, both in terms of richness and in terms of content features. But is it possible to measure the temporal dynamics of music-evoked visual imagery? If so, does visual imagery occur independent of time, or is it mostly guided by the structural and temporal dynamics of music? It is well understood that music affords the formation of narrative sequences (Dahl et al., 2022; Margulis, 2017; Margulis, Wong, et al., 2022; McAuley, Wong, Bellaiche, et al., 2021), which has the potential to vary as a function of musical genre (e.g., film music in Study 1, Chapter 2, of this thesis; instrumental music in Groves et al., 2023; and ambient music in Markert & Küssner, 2021). In order to capture changes in visual imagery over time, Presicce and Bailes (2019) collected continuous ratings of listeners' engagement with music and visual imagery occurrence in response to solo piano works and, using time series analyses, found the two experiences to be significantly associated. In a similar vein, Margulis, Williams et al. (2022) examined the relationship between perceived narratives and perceived tension by collecting responses over the course of a piece across multiple listenings of it, and found noticeable interindividual consistency on *when* a story event occurred which in turn coincided with perceived changes in musical tension. In a recently developed conceptual framework by Margulis and Jakubowski (2024) on the factors that underlie the generation of music-evoked autobiographical memories and fictional narratives, it is suggested that personal factors (based on prior made semantic, contextual, and emotional associations) and musical cues interact to shape the unfolding narratives that a person forms during music listening.

Future investigations guided by this framework and prior insights that, for instance, incorporate continuous ratings of music-evoked visual imagery alongside neuroscientific approaches, could allude more definitively to how discrete events in visual imagery unfold over time and are shaped by salient changes in the music.

7.4.3. Individual and situational differences

The present research has highlighted the need to consider an individual's personal traits in explaining their proclivities for forming music-evoked visual imagery, as well as the situational factors that guide their everyday imagery to music. Previous studies have typically aimed to understand each individual's ability to imagine in the first instance, as this has been shown to be linked with one's ability to recall information from pictures (Rodway et al., 2006). Furthermore, despite Studies 3 and 4 (Chapter 4 and 5) suggesting that most music-evoked visual imagery appear to listeners unbidden, the extent to which listeners are able to control their visual imagery at will is still an open question with limited and inconclusive findings (Hashim et al., 2020). Thus, there are several open questions in this regard.

With regard to situational factors, no research to date has explored the everyday occurrences of music-evoked visual mental imagery. Study 5b in Chapter 6 was able to reveal that individuals with aphantasia, when compared to the general population, use music differently in their everyday lives, with even more nuanced results found when assessing aphantasic subgroups varying in their inability to imagine (for instance, in addition to the reduced reminiscence that aphantasics experience to music, pure aphantasics additionally reported experiencing less strong emotional experiences). Thus, it would be very fruitful for future work to consider employing an experience sampling technique, whereby daily probes are sent to individuals, in order to relate certain imagery experiences to different contexts.

There is some recent work that has alluded to the prominence of visual imagery as a representational format of mind wandering (Taruffi, 2021). However, given the multimodal experiences that music can lead to (Nanay, 2015, 2018), it would be logical to assume that reports will encompass most, if not all, modalities of imagery, amongst which it would be fruitful to understand visual imagery's predominance and significance in relation to specific situations.

7.5. CLOSING STATEMENT

The present thesis sought to enhance understanding of the visual mental imagery that listeners form in their mind's eye when listening to music. Using a range of behavioural and neurophysiological methods, this research has reinforced previous findings and revealed several novel insights on the phenomenology of music-evoked visual imagery. It has helped to fill a gap into the neural signatures of music-evoked visual imagery varying in terms of content type as well as intentionality, providing evidence into its content and even helping to disentangle it from other forms of spontaneous cognition. Music-evoked visual imagery was revealed to be a functional tool, both as a potential technique in stress reduction and as a key representational format for music-evoked autobiographical memories. In all, the present research contributes to furthering our understanding of the extent to which music leads listeners to imagine and proposes numerous avenues for further research.

References

Abraham, A. (2016). The imaginative mind. *Human Brain Mapping*, *37*(11), 4197–4211. https://doi.org/10.1002/hbm.23300

Abraham, A. (2020). Surveying the Imagination Landscape. In A. Abraham (Ed.), *The Cambridge Handbook of the Imagination* (pp. 1–10). Cambridge University Press. https://doi.org/10.1017/9781108580298.001

Abraham, A., & Bubic, A. (2015). Semantic memory as the root of imagination. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00325

Aldworth, S. (2018). The art of imagination. *Cortex*, *105*, 173–181. https://doi.org/10.1016/j.cortex.2018.03.014

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. https://doi.org/10.1016/S0028-3932(00)00079-8

Alluri, V., Toiviainen, P., Lund, T. E., Wallentin, M., Vuust, P., Nandi, A. K., Ristaniemi, T., & Brattico, E. (2013). From Vivaldi to Beatles and back: Predicting lateralized brain responses to music. *NeuroImage*, *83*, 627–636. https://doi.org/10.1016/j.neuroimage.2013.06.064

Altenmüller, E., Schürmann, K., Lim, V. K., & Parlitz, D. (2002). Hits to the left, flops to the right: Different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia*, *40*(13), 2242–2256. https://doi.org/10.1016/S0028-3932(02)00107-0

Andrade, J., Kavanagh, D., & Baddeley, A. (1997). Eye-movements and visual imagery: A working memory approach to the treatment of post-traumatic stress disorder. *British*

Journal of Clinical Psychology, *36*(2), 209–223. https://doi.org/10.1111/j.2044-8260.1997.tb01408.x

Andrade, J., May, J., Deeprose, C., Baugh, S.-J., & Ganis, G. (2014). Assessing vividness of mental imagery: The Plymouth Sensory Imagery Questionnaire. *British Journal of Psychology*, *105*(4), 547–563. https://doi.org/10.1111/bjop.12050

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., & Buckner, R. L. (2010). Evidence for the Default Network's Role in Spontaneous Cognition. *Journal of Neurophysiology*, *104*(1), 322–335. https://doi.org/10.1152/jn.00830.2009

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, *65*(4), 550–562. https://doi.org/10.1016/j.neuron.2010.02.005

Antović, M., Küssner, M. B., Kempf, A., Omigie, D., Hashim, S., & Schiavio, A. (2024). "'A huge man is bursting out of a rock'": Bodies, motion, and creativity in verbal reports of musical connotation. *Journal of New Music Research*, *52*(1), 73–86. https://doi.org/10.1080/09298215.2024.2306406

Arcangeli, M. (2020). The Two Faces of Mental Imagery. *Philosophy and Phenomenological Research*, *101*(2), 304–322. https://doi.org/10.1111/phpr.12589

Arcangeli, M. (2023). Aphantasia demystified. *Synthese*, *201*(2), 31. https://doi.org/10.1007/s11229-022-04027-9

Baltazar, M., & Västfjäll, D. (2020). Songs Perceived as Relaxing: Musical Features, Lyrics, and Contributing Mechanisms. *PAM-IE 2019: Proceedings of the First International Conference: Psychology and Music – Interdisciplinary Encounters*, 115–124.

Baltazar, M., Västfjäll, D., Asutay, E., Koppel, L., & Saarikallio, S. (2019). Is it me or the music? Stress reduction and the role of regulation strategies and music. *Music & Science*, *2*, 205920431984416. https://doi.org/10.1177/2059204319844161

Balteș, F. R., & Miu, A. C. (2014). Emotions during live music performance: Links with individual differences in empathy, visual imagery, and mood. *Psychomusicology: Music, Mind, and Brain*, *24*(1), 58–65. https://doi.org/10.1037/pmu0000030

Bangert, M., & Altenmüller, E. O. (2003). Mapping perception to action in piano practice: A longitudinal DC-EEG study. *BMC Neuroscience*, *4*(1), 26. https://doi.org/10.1186/1471-2202-4-26

Barraza, P., Chavez, M., & Rodríguez, E. (2016). Ways of making-sense: Local gamma synchronization reveals differences between semantic processing induced by music and language. *Brain and Language*, *152*, 44–49. https://doi.org/10.1016/j.bandl.2015.12.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Belfi, A. M. (2019). Emotional valence and vividness of imagery predict aesthetic appeal in music. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 128–135. https://doi.org/10.1037/pmu0000232

Belfi, A. M. (2022). Neuroscience Measures of Music and Mental Imagery. In M. B. Küssner, L. Taruffi, & G. A. Floridou (Eds.), *Music and Mental Imagery* (pp. 101–111). Routledge.

Belfi, A. M., Vessel, E. A., & Starr, G. G. (2017). Individual Ratings of Vividness Predict Aesthetic Appeal in Poetry. *Psychology of Aesthetics, Creativity, and the Arts*. http://dx.doi.org/10.1037/aca0000153

Benedek, M., Bergner, S., Könen, T., Fink, A., & Neubauer, A. C. (2011). EEG alpha synchronization is related to top-down processing in convergent and divergent thinking.

Neuropsychologia, *49*(12), 3505–3511. https://doi.org/10.1016/j.neuropsychologia.2011.09.004

Bernardi, N. F., De Buglio, M., Trimarchi, P. D., Chielli, A., & Bricolo, E. (2013). Mental practice promotes motor anticipation: Evidence from skilled music performance. *Frontiers in Human Neuroscience*, *7*(451). https://doi.org/10.3389/fnhum.2013.00451

Bishop, L., Bailes, F., & Dean, R. T. (2013). Musical Imagery and the Planning of Dynamics and Articulation During Performance. *Music Perception*, *31*(2), 97–117. https://doi.org/10.1525/mp.2013.31.2.97

Bishop, S. R. (2002). What Do We Really Know About Mindfulness-Based Stress Reduction? *Psychosomatic Medicine*, *64*(1), 71.

Black, M. T. (2022). “Don’t Sing It With the Face of a Dead Fish!” Can Verbalized Imagery Stimulate a Vocal Response in Choral Rehearsals? In *Music and Mental Imagery*. Routledge.

Blackwell, S. E. (2020). Emotional Mental Imagery. In A. Abraham (Ed.), *The Cambridge Handbook of the Imagination* (pp. 241–257). Cambridge University Press. https://doi.org/10.1017/9781108580298.016

Bonde, L. O. (2007). Imagery, Metaphor, and Perceived Outcome in Six Cancer Survivors’ Bonny Method of Guided Imagery and Music (BMGIM) Therapy. *Qualitative Inquiries in Music Therapy; Gilsum*, *3*, 132–164.

Bonny, H. L. (1983). Music listening for intensive coronary care units: A pilot project. *Music Therapy*, *3*(1), 4–16. https://doi.org/10.1093/mt/3.1.4

Bonny, Helen. L. (1986). Music and Healing. *Music Therapy*, *6*(1), 3–12. https://doi.org/10.1093/mt/6.1.3

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes. *Memory & Cognition*, *36*(4), 849–862. https://doi.org/10.3758/MC.36.4.849

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: Neural markers of low alertness during mind wandering. *NeuroImage*, *54*(4), 3040–3047. https://doi.org/10.1016/j.neuroimage.2010.10.008

Bradley, M. M., & Lang, P. J. (2007). *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual.* [Technical report B-3.].

Braun, D. I., & Doerschner, K. (2019). Kandinsky or Me? How Free Is the Eye of the Beholder in Abstract Art? *I-Perception*, *10*(5), 204166951986797. https://doi.org/10.1177/2041669519867973

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bremner, J. D., Krystal, J. H., Southwick, S. M., & Charney, D. S. (1996). Noradrenergic mechanisms in stress and anxiety: I. preclinical studies. *Synapse*, *23*(1), 28–38. https://doi.org/10.1002/(SICI)1098-2396(199605)23:1<28::AID-SYN4>3.0.CO;2-J

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, *54*(2), 103–109. https://doi.org/10.1016/S0278-2626(03)00264-1

Cacha, L. A., Poznanski, R. R., Salleh, S. H., Latif, A. Z. A., & Ariff, T. M. (2020). Anxiety-related circuitry in affective neuroscience. *Journal of Molecular and Clinical Medicine*, *3*(3), Article 3. https://doi.org/10.31083/j.jmcm.2020.03.806

Carrus, E., Pearce, M. T., & Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex*, *49*(8), 2186–2200. https://doi.org/10.1016/j.cortex.2012.08.024

Cespedes-Guevara, J., & Dibben, N. (2022). The Role of Embodied Simulation and Visual Imagery in Emotional Contagion with Music. *Music & Science*, *5*, 20592043221093836. https://doi.org/10.1177/20592043221093836

Chafin, S., Roy, M., Gerin, W., & Christenfeld, N. (2004). Music can facilitate blood pressure recovery from stress. *British Journal of Health Psychology*, *9*(3), 393–403. https://doi.org/10.1348/1359107041557020

Chan, M. F., Wong, Z. Y., & Thayala, N. V. (2011). The effectiveness of music listening in reducing depressive symptoms in adults: A systematic review. *Complementary Therapies in Medicine*, *19*(6), 332–348. https://doi.org/10.1016/j.ctim.2011.08.003

Chen, G., Mishra, V., & Chen, C.-H. (2019). Temporal factors of listening to music on stress reduction. *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 907–914. https://doi.org/10.1145/3341162.3346272

Chen, N., Tanaka, K., & Watanabe, K. (2015). Color-Shape Associations Revealed with Implicit Association Tests. *PLOS ONE*, *10*(1), e0116954. https://doi.org/10.1371/journal.pone.0116954

Christian, B. M., Miles, L. K., Parkinson, C., & Macrae, C. N. (2013). Visual perspective and the characteristics of mind wandering. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00699

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*(21), 8719–8724. https://doi.org/10.1073/pnas.0900234106

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*(11), 718–731. https://doi.org/10.1038/nrn.2016.113

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

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

Compton, R. J., Gearinger, D., & Wild, H. (2019). The wandering mind oscillates: EEG alpha power is enhanced during moments of mind-wandering. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(5), 1184–1191. https://doi.org/10.3758/s13415-019-00745-9

Cooper, N. R., Croft, R. J., Dominey, S. J. J., Burgess, A. P., & Gruzelier, J. H. (2003). Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International Journal of Psychophysiology*, *47*(1), 65–74. https://doi.org/10.1016/S0167-8760(02)00107-1

Cross, I. (2001). Music, Cognition, Culture, and Evolution. *Annals of the New York Academy of Sciences*, *930*(1), 28–42. https://doi.org/10.1111/j.1749-6632.2001.tb05723.x

Cross, I. (2011). Music as a social and cognitive process. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 315–328). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199553426.003.0033

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474–478. https://doi.org/10.1016/j.visres.2006.11.013

Dahl, S., Stella, A., & Bjørner, T. (2022). Tell me what you see: An exploratory investigation of visual mental imagery evoked by music. *Musicae Scientiae*, 10298649221124862. https://doi.org/10.1177/10298649221124862

D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition*, *13*(4), 844–858. https://doi.org/10.1016/j.concog.2004.07.007

D'Argembeau, A., & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and Cognition*, *15*(2), 342–350. https://doi.org/10.1016/j.concog.2005.09.001

Davis, W. B., & Thaut, M. H. (1989). The Influence of Preferred Relaxing Music on Measures of State Anxiety, Relaxation, and Physiological Responses. *Journal of Music Therapy*, *26*(4), 168–187. https://doi.org/10.1093/jmt/26.4.168

Daviu, N., Bruchas, M. R., Moghaddam, B., Sandi, C., & Beyeler, A. (2019). Neurobiological links between stress and anxiety. *Neurobiology of Stress*, *11*, 100191. https://doi.org/10.1016/j.ynstr.2019.100191

Dawes, A. J., Keogh, R., Andrillon, T., & Pearson, J. (2020). A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific Reports*, *10*(1), 10022. https://doi.org/10.1038/s41598-020-65705-7

Dawes, A. J., Keogh, R., Robuck, S., & Pearson, J. (2022). Memories with a blind mind: Remembering the past and imagining the future with aphantasia. *Cognition*, *227*, 105192. https://doi.org/10.1016/j.cognition.2022.105192

Day, R. A., & Thompson, W. F. (2019). Measuring the onset of experiences of emotion and imagery in response to music. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 75–89. https://doi.org/10.1037/pmu0000220

Day, R. A., Thompson, W. F., & Boag, S. (2020). Characterizing experiences of music-evoked visual imagery in high prevalence contexts. *Psychomusicology: Music, Mind, and Brain*, *30*(2), 72–87. https://doi.org/10.1037/pmu0000251

de Borst, A. W., Sack, A. T., Jansma, B. M., Esposito, F., de Martino, F., Valente, G., Roebroeck, A., di Salle, F., Goebel, R., & Formisano, E. (2012). Integration of "what" and "where" in frontal cortex during visual imagery of scenes. *NeuroImage*, *60*(1), 47–58. https://doi.org/10.1016/j.neuroimage.2011.12.005

De Lange, F., Jensen, O., Bauer, M., & Toni, I. (2008). Interactions between posterior gamma and frontal alpha/beta oscillations during imagined actions. *Frontiers in Human Neuroscience*, *2*. https://www.frontiersin.org/articles/10.3389/neuro.09.007.2008

Decker, J. T., Brown, J. L. C., Ashley, W., & Lipscomb, A. E. (2019). Mindfulness, meditation, and breathing exercises: Reduced anxiety for clients and self-care for social work interns. *Social Work With Groups*, *42*(4), 308–322. https://doi.org/10.1080/01609513.2019.1571763

Deil, J., Markert, N., Normand, P., Kammen, P., Küssner, M. B., & Taruffi, L. (2023). Mind-wandering during contemporary live music: An exploratory study. *Musicae Scientiae*, *27*(3), 616–636. https://doi.org/10.1177/10298649221103210

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Denora, T. (2010). Emotion as Social Emergence: Perspectives from Music Sociology. In P. N. Juslin (Ed.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 158–183). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0007

Deroy, O. (2020). Evocation: How Mental Imagery Spans Across the Senses. In A. Abraham (Ed.), *The Cambridge Handbook of the Imagination* (1st ed., pp. 276–290). Cambridge University Press. https://doi.org/10.1017/9781108580298.018

Deroy, O., & Spence, C. (2016). Crossmodal Correspondences: Four Challenges. *Multisensory Research*, *29*(1–3), 29–48. https://doi.org/10.1163/22134808-00002488

Devitt, A. L., Addis, D. R., & Schacter, D. L. (2017). Episodic and semantic content of memory and imagination: A multilevel analysis. *Memory & Cognition*, *45*(7), 1078–1094. https://doi.org/10.3758/s13421-017-0716-1

Di Bona, E. (2017). Listening to the Space of Music. *Rivista Di Estetica*, *66*, Article 66. https://doi.org/10.4000/estetica.3112

Di Nuovo, S. F., & Angelica, A. (2015). Musical skills and perceived vividness of imagery: Differences between musicians and untrained subjects. *Annali della facoltà di Scienze della formazione - Università degli studi di Catania*, *14*, 3–13. https://doi.org/10.4420/unict-asdf.14.2015.1

Dijkstra, N. (2024). *Uncovering the Role of the Early Visual Cortex in Visual Mental Imagery* (2024021684). Preprints. https://doi.org/10.20944/preprints202402.1684.v1

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *The Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. a. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Scientific Reports*, *7*(1), 5677. https://doi.org/10.1038/s41598-017-05888-8

Dreksler, N., & Spence, C. (2019). A Critical Analysis of Colour–Shape Correspondences: Examining the Replicability of Colour–Shape Associations. *I-Perception*, *10*(2), 204166951983404. https://doi.org/10.1177/2041669519834042

Drever, J. (1955). Some observations on the occipital alpha rhythm. *Quarterly Journal of Experimental Psychology*, *7*(2), 91–97. https://doi.org/10.1080/17470215508416679

Dukić, H. (2022). In Search of a Story: Guided Imagery and Music Therapy. In *Music and Mental Imagery*. Routledge.

Dukić, H., Parncutt, R., & Bunt, L. (2019). Narrative archetypes in the imagery of clients in Guided Imagery and music therapy sessions. *Psychology of Music*, 030573561985412. https://doi.org/10.1177/0305735619854122

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Eerola, T., Vuoskoski, J. K., & Kautiainen, H. (2016). Being Moved by Unfamiliar Sad Music Is Associated with High Empathy. *Frontiers in Psychology*, *7*(1176). https://doi.org/10.3389/fpsyg.2016.01176

Egeth, H. E., & Yantis, S. (1997). Visual Attention: Control, Representation, and Time Course. *Annual Review of Psychology*, *48*(1), 269–297. https://doi.org/10.1146/annurev.psych.48.1.269

Ehrhardt, N. M., Fietz, J., Kopf-Beck, J., Kappelmann, N., & Brem, A. (2021). Separating EEG correlates of stress: Cognitive effort, time pressure, and social-evaluative threat. *European Journal of Neuroscience*, ejn.15211. https://doi.org/10.1111/ejn.15211

Eitan, Z., & Granot, R. Y. (2006). How Music Moves: Musical Parameters and Listeners Images of Motion. *Music Perception: An Interdisciplinary Journal*, *23*(3), 221–248. https://doi.org/10.1525/mp.2006.23.3.221

Fachner, J. C. (2022). Recumbent Journeys Into Sound—Music, Imagery, and Altering States of Consciousness. In *Music and Mental Imagery*. Routledge.

Fachner, J. C., Maidhof, C., Grocke, D., Nygaard Pedersen, I., Trondalen, G., Tucek, G., & Bonde, L. O. (2019). “Telling me not to worry…” Hyperscanning and Neural Dynamics of Emotion Processing During Guided Imagery and Music. *Frontiers in Psychology*, *10*(1561). https://doi.org/10.3389/fpsyg.2019.01561

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, *18*(1), 245–272. https://doi.org/10.1016/0010-0277(84)90026-X

Faw, B. (2009). Conflicting Intuitions May Be Based On Differing Abilities: Evidence from Mental Imaging Research. *Journal of Consciousness Studies*, *16*(4), 45–68.

Filgueiras, A., Quintas Conde, E. F., & Hall, C. R. (2018). The neural basis of kinesthetic and visual imagery in sports: An ALE meta − analysis. *Brain Imaging and Behavior*, *12*(5), 1513–1523. https://doi.org/10.1007/s11682-017-9813-9

Finch, K., & Moscovitch, D. A. (2016). Imagery-Based Interventions for Music Performance Anxiety: An Integrative Review. *Medical Problems of Performing Artists*, *31*(4), 222–231. https://doi.org/10.21091/mppa.2016.4040

Floridou, G. A. (2022). The Chronicles of Musical Imagery as It Occurs Before, During, and After Music. In *Music and Mental Imagery*. Routledge.

Fox, E., & McKinney, C. (2016). The Bonny Method of Guided Imagery and Music for Music Therapy Interns: A Survey of Effects on Professional and Personal Growth. *Music Therapy Perspectives; Washington, D. C.*, *34*(1), 90–98.

Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00154

Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., & Zeman, A. (2018). The neural correlates of visual imagery vividness – An fMRI study and literature review. *Cortex*, *105*, 26–40. https://doi.org/10.1016/j.cortex.2017.09.014

Gaesser, B. (2013). Constructing Memory, Imagination, and Empathy: A Cognitive Neuroscience Perspective. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00576

Gale, A., Morris, P. E., Lucas, B., & Richardson, A. (1972). Types of Imagery and Imagery Types: An Eeg Study. *British Journal of Psychology*, *63*(4), 523–531. https://doi.org/10.1111/j.2044-8295.1972.tb01302.x

Galton, F. (1880). Statistics of Mental Imagery. *Mind*, *5*(19), 301–318.

Gelding, R. W., Day, R. A., & Thompson, W. F. (2022). Music-Evoked Imagery and Imagery for Music: Subjective and Behavioural Measures. In *Music and Mental Imagery*. Routledge.

Gerra, G., Zaimovic, A., Franchini, D., Palladino, M., Giucastro, G., Reali, N., Maestri, D., Caccavari, R., Delsignore, R., & Brambilla, F. (1998). Neuroendocrine responses of healthy volunteers to `techno-music': Relationships with personality traits and emotional state. *International Journal of Psychophysiology*, *28*(1), 99–111. https://doi.org/10.1016/S0167-8760(97)00071-8

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, *14*(5), 10.1038/nrn3476. https://doi.org/10.1038/nrn3476

Grandchamp, R., & Delorme, A. (2011). Single-Trial Normalization for Event-Related Spectral Decomposition Reduces Sensitivity to Noisy Trials. *Frontiers in Psychology*, *2*. https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00236

Greenberg, D. L., & Knowlton, B. J. (2014). The role of visual imagery in autobiographical memory. *Memory & Cognition*, *42*(6), 922–934. https://doi.org/10.3758/s13421-014-0402-5

Gregory, D. (2016). Imagination and mental imagery. In *The Routledge Handbook of Philosophy of Imagination*. Routledge.

Groarke, J. M., Groarke, A., Hogan, M. J., Costello, L., & Lynch, D. (2020). Does Listening to Music Regulate Negative Affect in a Stressful Situation? Examining the Effects of Self-Selected and Researcher-Selected Music Using Both Silent and Active Controls. *Applied Psychology: Health and Well-Being*, *12*(2), 288–311. https://doi.org/10.1111/aphw.12185

Groarke, J. M., & Hogan, M. J. (2018). Development and Psychometric Evaluation of the Adaptive Functions of Music Listening Scale. *Frontiers in Psychology*, *9*, 516. https://doi.org/10.3389/fpsyg.2018.00516

Grocke, D. E. (2010). An Overview of Research in the Bonny Method of Guided Imagery and Music. *Voices: A World Forum for Music Therapy*, *10*(3). https://doi.org/10.15845/voices.v10i3.340

Grossman, P., Niemann, L., Schmidt, S., & Walach, H. (2004). Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of Psychosomatic Research*, *57*(1), 35–43. https://doi.org/10.1016/S0022-3999(03)00573-7

Groves, K., Farbood, M. M., Ripolles, P., & Zuanazzi, A. (2023). *Through the lens of music: Imagining movie scenes through soundtrack listening* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/aj97z

Gualberto Cremades, J. (2002). The effects of imagery perspective as a function of skill level on alpha activity. *International Journal of Psychophysiology*, *43*(3), 261–271. https://doi.org/10.1016/S0167-8760(01)00186-6

Han, C., Shapley, R., & Xing, D. (2022). Gamma rhythms in the visual cortex: Functions and mechanisms. *Cognitive Neurodynamics*, *16*(4), 745–756. https://doi.org/10.1007/s11571-021-09767-x

Handy, T., Miller, M., Schott, B., Shroff, N., Janata, P., Horn, J. V., Inati, S., Grafton, S., & Gazzaniga, M. (2004). Visual imagery and memory: Do retrieval strategies affect what the mind's eye sees? *European Journal of Cognitive Psychology*, *16*(5), 631–652. https://doi.org/10.1080/09541440340000457

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience & Biobehavioral Reviews*, *94*, 31–44. https://doi.org/10.1016/j.neubiorev.2018.08.003

Hargreaves, D. J. (2012). Musical imagination: Perception and production, beauty and creativity. *Psychology of Music*, *40*(5), 539–557. https://doi.org/10.1177/0305735612444893

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnint.2013.00083

Harmony, T., Fernández, T., Silva, J., Bernal, J., Díaz-Comas, L., Reyes, A., Marosi, E., Rodríguez, M., & Rodríguez, M. (1996). EEG delta activity: An indicator of attention to internal processing during performance of mental tasks. *International Journal of Psychophysiology*, *24*(1), 161–171. https://doi.org/10.1016/S0167-8760(96)00053-0

Harney, C., Johnson, J., Bailes, F., & Havelka, J. (2023). Is music listening an effective intervention for reducing anxiety? A systematic review and meta-analysis of controlled studies. *Musicae Scientiae*, *27*(2), 278–298. https://doi.org/10.1177/10298649211046979

Hashim, S., Stewart, L., & Küssner, M. B. (2020). Saccadic Eye-Movements Suppress Visual Mental Imagery and Partly Reduce Emotional Response During Music Listening. *Music & Science*, *3*, 205920432095958. https://doi.org/10.1177/2059204320959580

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726–1731. https://doi.org/10.1073/pnas.0610561104

Helsing, M., Västfjäll, D., Bjälkebring, P., Juslin, P. N., & Hartig, T. (2016). An Experimental Field Study of the Effects of Listening to Self-Selected Music on Emotions, Stress, and Cortisol Levels. *Music and Medicine*, *8*(4), 187–198.

Herff, S. A., Cecchetti, G., Ericson, P., & Cano, E. (2023). *Solitary Silence and Social Sounds: Music influences mental imagery, inducing thoughts of social interactions* [Preprint]. Animal Behavior and Cognition. https://doi.org/10.1101/2023.06.22.546175

Herff, S. A., Cecchetti, G., Taruffi, L., & Déguernel, K. (2021). Music influences vividness and content of imagined journeys in a directed visual imagery task. *Scientific Reports*, *11*(1), 15990. https://doi.org/10.1038/s41598-021-95260-8

Herff, S. A., McConnell, S., Ji, J. L., & Prince, J. B. (2022). Eye Closure Interacts with Music to Influence Vividness and Content of Directed Imagery. *Music & Science*, *5*, 20592043221142711. https://doi.org/10.1177/20592043221142711

Holmes, E. A., Brewin, C. R., & Hennessy, R. G. (2004). Trauma Films, Information Processing, and Intrusive Memory Development. *Journal of Experimental Psychology: General*, *133*(1), 3–22. https://doi.org/10.1037/0096-3445.133.1.3

Holmes, E. A., Geddes, J. R., Colom, F., & Goodwin, G. M. (2008). Mental imagery as an emotional amplifier: Application to bipolar disorder. *Behaviour Research and Therapy*, *46*(12), 1251–1258. https://doi.org/10.1016/j.brat.2008.09.005

Holmes, E. A., & Mathews, A. (2005). Mental Imagery and Emotion: A Special Relationship? *Emotion*, *5*(4), 489–497. https://doi.org/10.1037/1528-3542.5.4.489

Hölzel, B. K., Carmody, J., Evans, K. C., Hoge, E. A., Dusek, J. A., Morgan, L., Pitman, R. K., & Lazar, S. W. (2010). Stress reduction correlates with structural changes in the amygdala. *Social Cognitive and Affective Neuroscience*, *5*(1), 11–17. https://doi.org/10.1093/scan/nsp034

Irving, Z. C. (2016). Mind-wandering is unguided attention: Accounting for the “purposeful” wanderer. *Philosophical Studies*, *173*(2), 547–571. https://doi.org/10.1007/s11098-015-0506-1

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed Neural Systems for the Generation of Visual Images. *Neuron*, *28*(3), 979–990. https://doi.org/10.1016/S0896-6273(00)00168-9

Jakubowski, K. (2022). Mental Imagery in Music-Evoked Autobiographical Memories. In *Music and Mental Imagery*. Routledge.

Jakubowski, K., Belfi, A. M., & Eerola, T. (2021). Phenomenological Differences in Music- and Television-Evoked Autobiographical Memories. *Music Perception*, *38*(5), 435–455. https://doi.org/10.1525/mp.2021.38.5.435

Jakubowski, K., & Francini, E. (2022). Differential effects of familiarity and emotional expression of musical cues on autobiographical memory properties. *Quarterly Journal of Experimental Psychology*, 174702182211297. https://doi.org/10.1177/17470218221129793

Jakubowski, K., & Ghosh, A. (2021). Music-evoked autobiographical memories in everyday life. *Psychology of Music*, *49*(3), 649–666. https://doi.org/10.1177/0305735619888803

Jerath, R., Crawford, M. W., Barnes, V. A., & Harden, K. (2015). Self-Regulation of Breathing as a Primary Treatment for Anxiety. *Applied Psychophysiology and Biofeedback*, *40*(2), 107–115. https://doi.org/10.1007/s10484-015-9279-8

Johnson, M. L., & Larson, S. (2003). 'Something in the Way She Moves'—Metaphors of Musical Motion. *Metaphor and Symbol*, *18*(2), 63–84. https://doi.org/10.1207/S15327868MS1802_1

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235–266. https://doi.org/10.1016/j.plrev.2013.05.008

Juslin, P. N. (2019). *Musical Emotions Explained: Unlocking the Secrets of Musical Affect*. Oxford University Press.

Juslin, P. N., Barradas, G., & Eerola, T. (2015). From Sound to Significance: Exploring the Mechanisms Underlying Emotional Reactions to Music. *The American Journal of Psychology*, *128*(3), 281–304. https://doi.org/10.5406/amerjpsyc.128.3.0281

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*(4), 599–623. https://doi.org/10.1177/0305735613484548

Juslin, P. N., & Laukka, P. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*, *33*(3), 217–238. https://doi.org/10.1080/0929821042000317813

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion*, *8*(5), 668–683. https://doi.org/10.1037/a0013505

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575. https://doi.org/10.1017/S0140525X08005293

Kam, J. W. Y., Rahnuma, T., Park, Y. E., & Hart, C. M. (2022). Electrophysiological markers of mind wandering: A systematic review. *NeuroImage*, *258*, 119372. https://doi.org/10.1016/j.neuroimage.2022.119372

Kane, M. J., & McVay, J. C. (2012). What Mind Wandering Reveals About Executive-Control Abilities and Failures. *Current Directions in Psychological Science*, *21*(5), 348–354. https://doi.org/10.1177/0963721412454875

Kaufman, L., Schwartz, B., Salustri, C., & Williamson, S. J. (1990). Modulation of Spontaneous Brain Activity during Mental Imagery. *Journal of Cognitive Neuroscience*, *2*(2), 124–132. https://doi.org/10.1162/jocn.1990.2.2.124

Kavanagh, D. J., Freese, S., Andrade, J., & May, J. (2001). Effects of visuospatial tasks on desensitization to emotive memories. *British Journal of Clinical Psychology*, *40*(3), 267–280.

Kay, L., Keogh, R., Andrillon, T., & Pearson, J. (2022). The pupillary light response as a physiological index of aphantasia, sensory and phenomenological imagery strength. *eLife*, *11*, e72484. https://doi.org/10.7554/eLife.72484

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. https://doi.org/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*, *28*(1), 27–46. https://doi.org/10.1525/mp.2010.28.1.27

Kemp, A. H., & Felmingham, K. L. (2008). The psychology and neuroscience of depression and anxiety: Towards an integrative model of emotion disorders. *Psychology & Neuroscience*, *1*, 177–181. https://doi.org/10.3922/j.psns.2008.2.010

Kemps, E., & Tiggemann, M. (2007). Reducing the vividness and emotional impact of distressing autobiographical memories: The importance of modality-specific interference. *Memory*, *15*(4), 412–422. https://doi.org/10.1080/09658210701262017

Keogh, R., & Pearson, J. (2011). Mental Imagery and Visual Working Memory. *PLOS ONE*, *6*(12), e29221. https://doi.org/10.1371/journal.pone.0029221

Keogh, R., Pearson, J., & Zeman, A. (2021). Aphantasia: The science of visual imagery extremes. In *Handbook of Clinical Neurology* (Vol. 178, pp. 277–296). Elsevier. https://doi.org/10.1016/B978-0-12-821377-3.00012-X

Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex*, *143*, 237–253. https://doi.org/10.1016/j.cortex.2021.07.012

Klein, C., Liem, F., Hänggi, J., Elmer, S., & Jäncke, L. (2015). The “silent” imprint of musical training. *Human Brain Mapping*, *37*(2), 536–546. https://doi.org/10.1002/hbm.23045

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003

Knight, K. F., Milton, F., & Zeman, A. Z. J. (2022). Memory without Imagery: No Evidence of Visual Working Memory Impairment in People with Aphantasia. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/0b16s06v

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179–190. https://doi.org/10.1111/j.1469-8986.2008.00752.x

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180. https://doi.org/10.1038/nrn3666

Koelsch, S., Bashevkin, T., Kristensen, J., Tvedt, J., & Jentschke, S. (2019). Heroic music stimulates empowering thoughts during mind-wandering. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-46266-w

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between Syntax Processing in Language and in Music: An ERP Study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577. https://doi.org/10.1162/089892905774597290

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), Article 3. https://doi.org/10.1038/nn1197

Koelsch, S., Maess, B., Grossmann, T., & Friederici, A. D. (2003). Electric brain responses reveal gender differences in music processing. *NeuroReport*, *14*(5), 709–713. https://doi.org/10.1097/01.wnr.0000065762.60383.67

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage*, *81*, 49–60. https://doi.org/10.1016/j.neuroimage.2013.05.008

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179088.001.0001

Kramer, J. H., Delis, D. C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology*, *44*(6), 907–915. https://doi.org/10.1002/1097-4679(198811)44:6<907::AID-JCLP2270440610>3.0.CO;2-8

Kraus, N., & Slater, J. (2015). Chapter 12 - Music and language: Relations and disconnections. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 207–222). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00012-3

Küssner, M. B., & Eerola, T. (2017). *The special case of music-induced visual imagery and its correlates with musical skills: Findings from an online survey*. Conference of the European Society for the Cognitive Sciences of Music (ESCOM), Ghent, Belgium.

Küssner, M. B., & Eerola, T. (2019). The content and functions of vivid and soothing visual imagery during music listening: Findings from a survey study. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 90–99. https://doi.org/10.1037/pmu0000238

Küssner, M., Taruffi, L., & Floridou, G. A. (Eds.). (2023). *Music and mental imagery*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9780429330070

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Larson, R. (1995). Secrets in the bedroom: Adolescents' private use of media. *Journal of Youth and Adolescence*, *24*(5), 535–550. https://doi.org/10.1007/BF01537055

Lazarus, R. S., Speisman, J. C., & Mordkoff, A. M. (1963). The Relationship Between Autonomic Indicators of Psychological Stress: Heart Rate and Skin Conductance: *Psychosomatic Medicine*, *25*(1), 19–30. https://doi.org/10.1097/00006842-196301000-00004

Lee, K.-C., Chao, Y.-H., Yiin, J.-J., Hsieh, H.-Y., Dai, W.-J., & Chao, Y.-F. (2012). Evidence That Music Listening Reduces Preoperative Patients' Anxiety. *Biological Research For Nursing*, *14*(1), 78–84. https://doi.org/10.1177/1099800410396704

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Lehmann, D., Faber, P. L., Achermann, P., Jeanmonod, D., Gianotti, L. R. R., & Pizzagalli, D. (2001). Brain sources of EEG gamma frequency during volitionally meditation-induced, altered states of consciousness, and experience of the self. *Psychiatry Research: Neuroimaging*, *108*(2), 111–121. https://doi.org/10.1016/S0925-4927(01)00116-0

Lejuez, C. W., Kahler, C. W., & Brown, R. A. (2003). A modified computer version of the Paced Auditory Serial Addition Task (PASAT) as a laboratory-based stressor. *The Behavior Therapist*, *26*(4), 290–293.

Leman, M., Maes, P.-J., Nijs, L., & Van Dyck, E. (2018). What Is Embodied Music Cognition? In R. Bader (Ed.), *Springer Handbook of Systematic Musicology* (pp. 747–760). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-55004-5_34

Lenormand, D., Fauvel, B., & Piolino, P. (2024). The formation of episodic autobiographical memory is predicted by mental imagery, self-reference, and anticipated details. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1355343

Levinson, J. (2004). Music as Narrative and Music as Drama. *Mind and Language*, *19*(4), 428–441. https://doi.org/10.1111/j.0268-1064.2004.00267.x

Liu, Y., Zhao, J., Zhou, X., Liu, X., Chen, H., & Yuan, H. (2021). The Neural Markers of Self-Caught and Probe-Caught Mind Wandering: An ERP Study. *Brain Sciences*, *11*(10), Article 10. https://doi.org/10.3390/brainsci11101329

Lotze, M. (2013). Kinesthetic imagery of musical performance. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00280

Lotze, M., & Halsband, U. (2006). Motor imagery. *Journal of Physiology-Paris*, *99*(4), 386–395. https://doi.org/10.1016/j.jphysparis.2006.03.012

Luft, C. D. B., Zioga, I., Banissy, M. J., & Bhattacharya, J. (2019). Spontaneous Visual Imagery During Meditation for Creating Visual Art: An EEG and Brain Stimulation Case Study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00210

MacDonald, L. E. (2013). *The Invisible Art of Film Music: A Comprehensive History*. Scarecrow Press.

MacGregor, C., & Müllensiefen, D. (2019). The Musical Emotion Discrimination Task: A New Measure for Assessing the Ability to Discriminate Emotions in Music. *Frontiers in Psychology*, *10*, 1955. https://doi.org/10.3389/fpsyg.2019.01955

MacGregor, C., Ruth, N., & Müllensiefen, D. (2023). Development and validation of the first adaptive test of emotion perception in music. *Cognition and Emotion*, *37*(2), 284–302. https://doi.org/10.1080/02699931.2022.2162003

Maidhof, C., & Koelsch, S. (2011). Effects of Selective Attention on Syntax Processing in Music and Language. *Journal of Cognitive Neuroscience*, *23*(9), 2252–2267. https://doi.org/10.1162/jocn.2010.21542

Makin, A. D. J., & Wuerger, S. M. (2013). The IAT shows no evidence for Kandinsky's color-shape associations. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00616

Margulis, E. H. (2017). An Exploratory Study of Narrative Experiences of Music. *Music Perception: An Interdisciplinary Journal*, *35*(2), 235–248. https://doi.org/10.1525/mp.2017.35.2.235

Margulis, E. H., & Jakubowski, K. (2024). Music, Memory, and Imagination. *Current Directions in Psychological Science*, 09637214231217229. https://doi.org/10.1177/09637214231217229

Margulis, E. H., Williams, J., Simchy-Gross, R., & McAuley, J. D. (2022). When did that happen? The dynamic unfolding of perceived musical narrative. *Cognition*, *226*, 105180. https://doi.org/10.1016/j.cognition.2022.105180

Margulis, E. H., Wong, P. C. M., Turnbull, C., Kubit, B. M., & McAuley, J. D. (2022). Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity. *Proceedings of the National Academy of Sciences*, *119*(4), e2110406119. https://doi.org/10.1073/pnas.2110406119

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markert, N., & Küssner, M. B. (2021). An exploratory study of visual mental imagery induced by ambient music. In J. Stupacher & S. Hagner (Eds.), *Proceedings of the 14th International Conference of Students of Systematic Musicology (SysMus21)*. Open Science Framework. https://doi.org/10.17605/OSF.IO/HG6RZ

Marks, D. F. (1973). Visual Imagery Differences in the Recall of Pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Martarelli, C. S., Mayer, B., & Mast, F. W. (2016). Daydreams and trait affect: The role of the listener's state of mind in the emotional response to music. *Consciousness and Cognition*, *46*, 27–35. https://doi.org/10.1016/j.concog.2016.09.014

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Marti-Marca, A., Nguyen, T., & Grahn, J. A. (2020). Keep Calm and Pump Up the Jams: How Musical Mood and Arousal Affect Visual Attention. *Music & Science*, *3*, 205920432092273. https://doi.org/10.1177/2059204320922737

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual Differences in Music Reward Experiences. *Music Perception*, *31*(2), 118–138. https://doi.org/10.1525/mp.2013.31.2.118

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Maus, F. E. (1991). Music As Narrative. *Indiana Theory Review*, *12*, 1–34.

McAuley, J. D., Wong, P. C. M., Bellaiche, L., & Margulis, E. H. (2021). What Drives Narrative Engagement With Music? *Music Perception*, *38*(5), 509–521. https://doi.org/10.1525/mp.2021.38.5.509

McAuley, J. D., Wong, P. C. M., Mamidipaka, A., Phillips, N., & Margulis, E. H. (2021). Do you hear what I hear? Perceived narrative constitutes a semantic dimension for music. *Cognition*, *212*, 104712. https://doi.org/10.1016/j.cognition.2021.104712

McClary, S. (1993). Narrative Agendas in 'Absolute' Music: Identity and Difference in Brahms's Third Symphony. In R. A. Solie (Ed.), *Musicology and Difference* (pp. 326–344). University of California Press. https://doi.org/10.1525/9780520916500-017

McKelvie, S. J., & Demers, E. G. (1979). Individual differences in reported visual imagery and memory performance. *British Journal of Psychology*, *70*(1), 51–57.

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. *Cerebral Cortex*, *14*(11), 1256–1265. https://doi.org/10.1093/cercor/bhh087

Menicucci, D., Di Gruttola, F., Cesari, V., Gemignani, A., Manzoni, D., & Sebastiani, L. (2020). Task-independent Electrophysiological Correlates of Motor Imagery Ability from Kinaesthetic and Visual Perspectives. *Neuroscience*, *443*, 176–187. https://doi.org/10.1016/j.neuroscience.2020.07.038

Minguillon, J., Lopez-Gordo, M. A., & Pelayo, F. (2016). Stress Assessment by Prefrontal Relative Gamma. *Frontiers in Computational Neuroscience*, *10*. https://doi.org/10.3389/fncom.2016.00101

Monzel, M., Vetterlein, A., & Reuter, M. (2023). No general pathological significance of aphantasia: An evaluation based on criteria for mental disorders. *Scandinavian Journal of Psychology*, *64*(3), 314–324. https://doi.org/10.1111/sjop.12887

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1273–1280. https://doi.org/10.1098/rstb.2008.0314

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089642

Mutha, P. K., Stapp, L. H., Sainburg, R. L., & Haaland, K. Y. (2014). Frontal and parietal cortex contributions to action modification. *Cortex*, *57*, 38–50. https://doi.org/10.1016/j.cortex.2014.03.005

Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies*, *172*(7), 1723–1736.

Nanay, B. (2018). Multimodal mental imagery. *Cortex*, *105*, 125–134. https://doi.org/10.1016/j.cortex.2017.07.006

Nanay, B. (2021). Mental Imagery. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/mental-imagery/

Nattiez, J.-J. (1990). Can One Speak of Narrativity in Music? *Journal of the Royal Musical Association*, *115*(2), 240–257. https://doi.org/10.1093/jrma/115.2.240

Neuper, C., & Pfurtscheller, G. (1996). Post-movement synchronization of beta rhythms in the EEG over the cortical foot area in man. *Neuroscience Letters*, *216*(1), 17–20. https://doi.org/10.1016/0304-3940(96)12991-8

Newcomb, A. (1987). Schumann and Late Eighteenth-Century Narrative Strategies. *19th-Century Music*, *11*(2), 164–174. https://doi.org/10.2307/746729

Nishizaka, A. (2003). Imagination in Action. *Theory & Psychology*, *13*(2), 177–207. https://doi.org/10.1177/0959354303013002002

Nolte, T., Guiney, J., Fonagy, P., Mayes, L. C., & Luyten, P. (2011). Interpersonal Stress Regulation and the Development of Anxiety Disorders: An Attachment-Based Developmental Framework. *Frontiers in Behavioral Neuroscience*, *5*. https://doi.org/10.3389/fnbeh.2011.00055

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of Music in Everyday Life. *Music Perception: An Interdisciplinary Journal*, *22*(1), 41–77. https://doi.org/10.1525/mp.2004.22.1.41

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Otenen, E., & Aydin, C. (2023). Exploring the Role of Visual Imagery in the Recall of Emotional Autobiographical Memories. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). https://escholarship.org/uc/item/4cv1c15q

Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscillations. *Trends in Neurosciences*, *30*(4), 150–158. https://doi.org/10.1016/j.tins.2007.02.001

Panteleeva, Y., Ceschi, G., Glowinski, D., Courvoisier, D. S., & Grandjean, D. (2018). Music for anxiety? Meta-analysis of anxiety reduction in non-clinical samples. *Psychology of Music*, *46*(4), 473–487. https://doi.org/10.1177/0305735617712424

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), Article 10. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J. (2020). The Visual Imagination. In A. Abraham (Ed.), *The Cambridge Handbook of the Imagination* (pp. 175–186). Cambridge University Press. https://doi.org/10.1017/9781108580298.012

Pearson, J., & Keogh, R. (2019). Redefining Visual Working Memory: A Cognitive-Strategy, Brain-Region Approach. *Current Directions in Psychological Science*, *28*(3), 266–273. https://doi.org/10.1177/0963721419835210

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. https://doi.org/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. https://doi.org/10.1016/j.tics.2015.08.003

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the Mind's Eye: The Metacognition of Visual Imagery. *Psychological Science*, *22*(12), 1535–1542. https://doi.org/10.1177/0956797611417134

Pearson, J., & Westbrook, F. (2015). Phantom perception: Voluntary and involuntary nonretinal vision. *Trends in Cognitive Sciences*, *19*(5), 278–284. https://doi.org/10.1016/j.tics.2015.03.004

Pekala, R. J. (1991). The Phenomenology of Consciousness Inventory. In *Quantifying Consciousness: Emotions, Personality, and Psychotherapy*. Springer.

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and Emotions in the Brain: Familiarity Matters. *PLOS ONE*, *6*(11), e27241. https://doi.org/10.1371/journal.pone.0027241

Petsche, H. (1996). Approaches to verbal, visual and musical creativity by EEG coherence analysis. *International Journal of Psychophysiology*, *24*(1–2), 145–159. https://doi.org/10.1016/S0167-8760(96)00050-5

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science*, *308*(5727), 1430–1430. https://doi.org/10.1126/science.1110922

Polychroni, N., Herrojo Ruiz, M., & Terhune, D. B. (2022). Introspection confidence predicts EEG decoding of self-generated thoughts and meta-awareness. *Human Brain Mapping*, *43*(7), 2311–2327. https://doi.org/10.1002/hbm.25789

Presicce, G. (2022). The Image Behind the Sound: Visual Imagery in Music Performance. In *Music and Mental Imagery*. Routledge.

Presicce, G., & Bailes, F. (2019). Engagement and visual imagery in music listening: An exploratory study. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 136–155. https://doi.org/10.1037/pmu0000243

R Core Team. (2018). *R: A Language and Environment for Statistical Computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Radstaak, M., Geurts, S. A. E., Brosschot, J. F., & Kompier, M. A. J. (2014). Music and Psychophysiological Recovery from Stress. *Psychosomatic Medicine*, *76*(7), 529. https://doi.org/10.1097/PSY.0000000000000094

Raffman, D. (1993). *Language, Music, and Mind*. The MIT Press.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor Cues During Slow-Wave Sleep Prompt Declarative Memory Consolidation. *Science*, *315*(5817), 1426–1429. https://doi.org/10.1126/science.1138581

Reinhardt, T., Schmahl, C., Wüst, S., & Bohus, M. (2012). Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim Multicomponent Stress Test (MMST). *Psychiatry Research*, *198*(1), 106–111. https://doi.org/10.1016/j.psychres.2011.12.009

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371–388. https://doi.org/10.1177/0305735604046096

Rodway, P., Gillies, K., & Schepman, A. (2006). Vivid imagers are better at detecting salient changes. *Journal of Individual Differences*, *27*(4), 218–228. https://doi.org/10.1027/1614-0001.27.4.218

Rosenberg, H. S., & Trusheim, W. (1989). Creative Transformations: How Visual Artists, Musicians, and Dancers Use Mental Imagery in Their Work. In J. E. Shorr, P. Robin, J. A. Connella, & M. Wolpin (Eds.), *Imagery: Current Perspectives* (pp. 55–75). Springer US. https://doi.org/10.1007/978-1-4899-0876-6_6

Rubin, D. C., & Umanath, S. (2015). Event Memory: A Theory of Memory for Laboratory, Autobiographical, and Fictional Events. *Psychological Review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74–85. https://doi.org/10.1037/0012-1649.37.1.74

Salenius, S., Kajola, M., Thompson, W. L., Kosslyn, S., & Hari, R. (1995). Reactivity of magnetic parieto-occipital alpha rhythm during visual imagery. *Electroencephalography and Clinical Neurophysiology*, *95*(6), 453–462. https://doi.org/10.1016/0013-4694(95)00155-7

Salmelin, R., Hámáaláinen, M., Kajola, M., & Hari, R. (1995). Functional Segregation of Movement-Related Rhythmic Activity in the Human Brain. *NeuroImage*, *2*(4), 237–243. https://doi.org/10.1006/nimg.1995.1031

Sauseng, P., Klimesch, W., Doppelmayr, M., Pecherstorfer, T., Freunberger, R., & Hanslmayr, S. (2005). EEG alpha synchronization and functional coupling during top-down processing in a working memory task. *Human Brain Mapping*, *26*(2), 148–155. https://doi.org/10.1002/hbm.20150

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Schaefer, R. S. (2022). Imagery and Movement in Music-Based Rehabilitation and Music Pedagogy. In *Music and Mental Imagery*. Routledge.

Schaerlaeken, S., Glowinski, D., Rappaz, M.-A., & Grandjean, D. (2019). “Hearing music as . . .”: Metaphors evoked by the sound of classical music. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 100–116. https://doi.org/10.1037/pmu0000233

Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L., & Schulz, P. (2011). The perceived stress reactivity scale: Measurement invariance, stability, and validity in three countries. *Psychological Assessment*, *23*(1), 80–94. https://doi.org/10.1037/a0021148

Schomer, D. L., & Silva, F. L. da. (2012). *Niedermeyer’s Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins.

Schuster, C., Hilfiker, R., Amft, O., Scheidhauer, A., Andrews, B., Butler, J., Kischka, U., & Ettlin, T. (2011). Best practice for motor imagery: A systematic literature review on motor imagery training elements in five different disciplines. *BMC Medicine*, *9*(1), 1–35. https://doi.org/10.1186/1741-7015-9-75

Seli, P., Jonker, T. R., Cheyne, J. A., Cortes, K., & Smilek, D. (2015). Can research participants comment authoritatively on the validity of their self-reports of mind wandering and task engagement? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 703–709. https://doi.org/10.1037/xhp0000029

Seli, P., Risko, E. F., & Smilek, D. (2016). On the Necessity of Distinguishing Between Unintentional and Intentional Mind Wandering. *Psychological Science*, *27*(5), 685–691. https://doi.org/10.1177/0956797616634068

Sepúlveda, M. L. A., Alonso, J. L., Guevara, M. A., & González, M. H. (2014). Increased Prefrontal-Parietal EEG Gamma Band Correlation during Motor Imagery in Expert Video Game Players. *Actualidades En Psicología*, *28*(117), Article 117. https://doi.org/10.15517/ap.v28i117.14095

Sirigu, A., & Duhamel, J. R. (2001). Motor and Visual Imagery as Two Complementary but Neurally Dissociable Mental Processes. *Journal of Cognitive Neuroscience*, *13*(7), 910–919. https://doi.org/10.1162/089892901753165827

Sousa, T., Amaral, C., Andrade, J., Pires, G., Nunes, U. J., & Castelo-Branco, M. (2017). Pure visual imagery as a potential approach to achieve three classes of control for implementation of BCI in non-motor disorders. *Journal of Neural Engineering*, *14*(4), 046026. https://doi.org/10.1088/1741-2552/aa70ac

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Steinbeis, N., & Koelsch, S. (2008a). Comparing the Processing of Music and Language Meaning Using EEG and fMRI Provides Evidence for Similar and Distinct Neural Representations. *PLOS ONE*, *3*(5), e2226. https://doi.org/10.1371/journal.pone.0002226

Steinbeis, N., & Koelsch, S. (2008b). Shared Neural Resources between Music and Language Indicate Semantic Processing of Musical Tension-Resolution Patterns. *Cerebral Cortex*, *18*(5), 1169–1178. https://doi.org/10.1093/cercor/bhm149

Sternin, A., McGarry, L. M., Owen, A. M., & Grahn, J. A. (2021). The Effect of Familiarity on Neural Representations of Music and Language. *Journal of Cognitive Neuroscience*, 1–17. https://doi.org/10.1162/jocn_a_01737

Stevens, C. J. (2012). Music Perception and Cognition: A Review of Recent Cross-Cultural Research. *Topics in Cognitive Science*, *4*(4), 653–667. https://doi.org/10.1111/j.1756-8765.2012.01215.x

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery. *The Journal of Neuroscience*, *29*(5), 1565–1572. https://doi.org/10.1523/JNEUROSCI.4657-08.2009

Stratton, V. N., & Zalanowski, A. H. (1992). The Interfering Effects of Music with Imagery. *Imagination, Cognition and Personality*, *11*(4), 381–388. https://doi.org/10.2190/WA5E-CR32-Q549-MX77

Talamini, F., Vigl, J., Doerr, E., Grassi, M., & Carretti, B. (2022). Auditory and visual mental imagery in musicians and non-musicians. *Musicae Scientiae*, 102986492110627. https://doi.org/10.1177/10298649211062724

Tan, X., Yowler, C. J., Super, D. M., & Fratianne, R. B. (2012). The Interplay of Preference, Familiarity and Psychophysical Properties in Defining Relaxation Music. *Journal of Music Therapy*, *49*(2), 150–179. https://doi.org/10.1093/jmt/49.2.150

Taruffi, L. (2021). Mind-Wandering during Personal Music Listening in Everyday Life: Music-Evoked Emotions Predict Thought Valence. *International Journal of Environmental Research and Public Health*, *18*(23), Article 23. https://doi.org/10.3390/ijerph182312321

Taruffi, L., & Koelsch, S. (2014). The Paradox of Music-Evoked Sadness: An Online Survey. *PLOS ONE*, *9*(10), e110490. https://doi.org/10.1371/journal.pone.0110490

Taruffi, L., & Küssner, M. B. (2019). A review of music-evoked visual mental imagery: Conceptual issues, relation to emotion, and functional outcome. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 62–74. https://doi.org/10.1037/pmu0000226

Taruffi, L., & Küssner, M. B. (2022). Visual mental imagery, music, and emotion: From academic discourse to clinical applications. In M. B. Küssner, L. Taruffi, & G. A. Floridou (Eds.), *Music and Mental Imagery* (1st ed., pp. 32–41). Routledge. https://doi.org/10.4324/9780429330070-4

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. https://doi.org/10.1038/s41598-017-14849-0

Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2018, July 23). *Sad Music, Empathy, and Visual Mental Imagery: An fMRI Study*. International Conference on Music Perception and Cognition (ICMPC) and Conference of the European Society for the Cognitive Sciences of Music (ESCOM), Graz, Austria. https://static.uni-graz.at/fileadmin/veranstaltungen/music-psychology-conference2018/documents/ICMPC15ESCOM10abstractbook.pdf

Taruffi, L., Skouras, S., Pehrs, C., & Koelsch, S. (2021). Trait Empathy Shapes Neural Responses Toward Sad Music. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-020-00861-x

Thoma, M. V., Marca, R. L., Brönnimann, R., Finkel, L., Ehlert, U., & Nater, U. M. (2013). The Effect of Music on the Human Stress Response. *PLOS ONE*, *8*(8), e70156. https://doi.org/10.1371/journal.pone.0070156

Thomas, N. J. T. (1997). Mental Imagery. In *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2021/entries/mental-imagery/

Thompson, W. L., Slotnick, S. D., Burrage, M. S., & Kosslyn, S. M. (2009). Two Forms of Spatial Imagery: Neuroimaging Evidence. *Psychological Science*, *20*(10), 1245–1253. https://doi.org/10.1111/j.1467-9280.2009.02440.x

Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. *Topics in Cognitive Science*, *4*(4), 568–584. https://doi.org/10.1111/j.1756-8765.2012.01209.x

Trainor, L. J., Shahin, A. J., & Roberts, L. E. (2009). Understanding the Benefits of Musical Training. *Annals of the New York Academy of Sciences*, *1169*(1), 133–142. https://doi.org/10.1111/j.1749-6632.2009.04589.x

van den Hout, M. A., Engelhard, I. M., Beetsma, D., Slofstra, C., Hornsveld, H., Houtveen, J., & Leer, A. (2011). EMDR and mindfulness. Eye movements and attentional breathing tax working memory and reduce vividness and emotionality of aversive ideation. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(4), 423–431. https://doi.org/10.1016/j.jbtep.2011.03.004

Vanhollebeke, G., De Smet, S., De Raedt, R., Baeken, C., van Mierlo, P., & Vanderhasselt, M.-A. (2022). The neural correlates of psychosocial stress: A systematic review and meta-analysis of spectral analysis EEG studies. *Neurobiology of Stress*, *18*, 100452. https://doi.org/10.1016/j.ynstr.2022.100452

Villena-González, M., Palacios-García, I., Rodríguez, E., & López, V. (2018). Beta Oscillations Distinguish Between Two Forms of Mental Imagery While Gamma and Theta Activity Reflects Auditory Attention. *Frontiers in Human Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnhum.2018.00389

Vredeveldt, A., Hitch, G. J., & Baddeley, A. D. (2011). Eyeclosure helps memory by reducing cognitive load and enhancing visualisation. *Memory & Cognition*, *39*(7), 1253–1263. https://doi.org/10.3758/s13421-011-0098-8

Vroegh, T. (2018, May 16). *Investigating the directional link between music-induced visual imagery and two different types of emotional responses*. Poster presented at the KOSMOS Workshop ‘Mind Wandering and Visual Mental Imagery in Music’, Berlin, Germany.

Vroegh, T. (2019). Zoning-in or tuning-in? Identifying distinct absorption states in response to music. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 156–170. https://doi.org/10.1037/pmu0000241

Vroegh, T. (2021). Visual imagery in the listener’s mind: A network analysis of absorbed consciousness: Psychology of Consciousness: Theory, Research, and Practice. *Psychology of Consciousness: Theory, Research, and Practice*. https://doi.org/10.1037/cns0000274

Vroegh, T. (2022). Music-Evoked Imagery in an Absorbed State of Mind: A Bayesian Network Approach. In *Music and Mental Imagery*. Routledge.

Vuoskoski, J. K., & Eerola, T. (2015). Extramusical information contributes to emotions induced by music. *Psychology of Music*, *43*(2), 262–274. https://doi.org/10.1177/0305735613502373

Vuoskoski, J. K., Thompson, W. F., McIlwain, D., & Eerola, T. (2012). Who Enjoys Listening to Sad Music and Why? *Music Perception; Berkeley*, *29*(3), 311–317.

Watson, J. B. (1928). *THE WAYS OF BEHAVIORISM.* (First Edition). Harper & Brothers.

Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods*, *50*(2), 642–661. https://doi.org/10.3758/s13428-017-0891-9

Wicken, M., Keogh, R., & Pearson, J. (2021). The critical role of mental imagery in human emotion: Insights from fear-based imagery and aphantasia. *Proceedings of the Royal*

Society B: Biological Sciences, *288*(1946), 20210267. https://doi.org/10.1098/rspb.2021.0267

Williamson, S. J., Kaufman, L., Lu, Z.-L., Wang, J.-Z., & Karron, D. (1997). Study of human occipital alpha rhythm: The alphon hypothesis and alpha suppression. *International Journal of Psychophysiology*, *26*(1), 63–76. https://doi.org/10.1016/S0167-8760(97)00756-3

Wilson, V. E., Dikman, Z., Bird, E. I., Williams, J. M., Harmison, R., Shaw-Thornton, L., & Schwartz, G. E. (2016). EEG Topographic Mapping of Visual and Kinesthetic Imagery in Swimmers. *Applied Psychophysiology and Biofeedback*, *41*(1), 121–127. https://doi.org/10.1007/s10484-015-9307-8

Wong, S. S. H., & Lim, S. W. H. (2017). Mental imagery boosts music compositional creativity. *PLOS ONE*, *12*(3), e0174009. https://doi.org/10.1371/journal.pone.0174009

Woody, A., Hooker, E. D., Zoccola, P. M., & Dickerson, S. S. (2018). Social-evaluative threat, cognitive load, and the cortisol and cardiovascular stress response. *Psychoneuroendocrinology*, *97*, 149–155. https://doi.org/10.1016/j.psyneuen.2018.07.009

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Current Biology*, *30*(13), 2621-2627.e5. https://doi.org/10.1016/j.cub.2020.04.074

Yehuda, N. (2011). Music and Stress. *Journal of Adult Development*, *18*(2), 85–94. https://doi.org/10.1007/s10804-010-9117-4

Yinger, O. S., & Gooding, L. F. (2015). A Systematic Review of Music-Based Interventions for Procedural Support. *Journal of Music Therapy*, *52*(1), 1–77. https://doi.org/10.1093/jmt/thv004

Zabielska-Mendyk, E., Francuz, P., Jaśkiewicz, M., & Augustynowicz, P. (2018). The Effects of Motor Expertise on Sensorimotor Rhythm Desynchronization during Execution and Imagery of Sequential Movements. *Neuroscience*, *384*, 101–110. https://doi.org/10.1016/j.neuroscience.2018.05.028

Zacks, J. M. (2008). Neuroimaging Studies of Mental Rotation: A Meta-analysis and Review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19. https://doi.org/10.1162/jocn.2008.20013

Zeman, A. (2020). Aphantasia. In A. Abraham (Ed.), *The Cambridge Handbook of the Imagination* (pp. 692–710). Cambridge University Press. https://doi.org/10.1017/9781108580298.042

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery – Congenital aphantasia. *Cortex*, *73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019

Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., Hattersley, A., Heuerman-Williamson, B., Jones, K., MacKisack, M., & Winlove, C. (2020). Phantasia–The psychological significance of lifelong visual imagery vividness extremes. *Cortex*, *130*, 426–440. https://doi.org/10.1016/j.cortex.2020.04.003

Zeman, A. Z. J., Della Sala, S., Torrens, L. A., Gountouna, V.-E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination'. *Neuropsychologia*, *48*(1), 145–155. https://doi.org/10.1016/j.neuropsychologia.2009.08.024

Zhou, L., Jiang, C., Wu, Y., & Yang, Y. (2015). Conveying the concept of movement in music: An event-related brain potential study. *Neuropsychologia*, *77*, 128–136. https://doi.org/10.1016/j.neuropsychologia.2015.07.029

Ziv, N., Rotem, T., Arnon, Z., & Haimov, I. (2008). The Effect of Music Relaxation versus Progressive Muscular Relaxation on Insomnia in Older People and Their Relationship to Personality Traits. *Journal of Music Therapy*, *45*(3), 360–380. https://doi.org/10.1093/jmt/45.3.360

Appendix A. Chapter 2 Supplementary Tables and Figures

Figure A.1.

Level 2 content of visual imagery frequencies divided by excerpt.

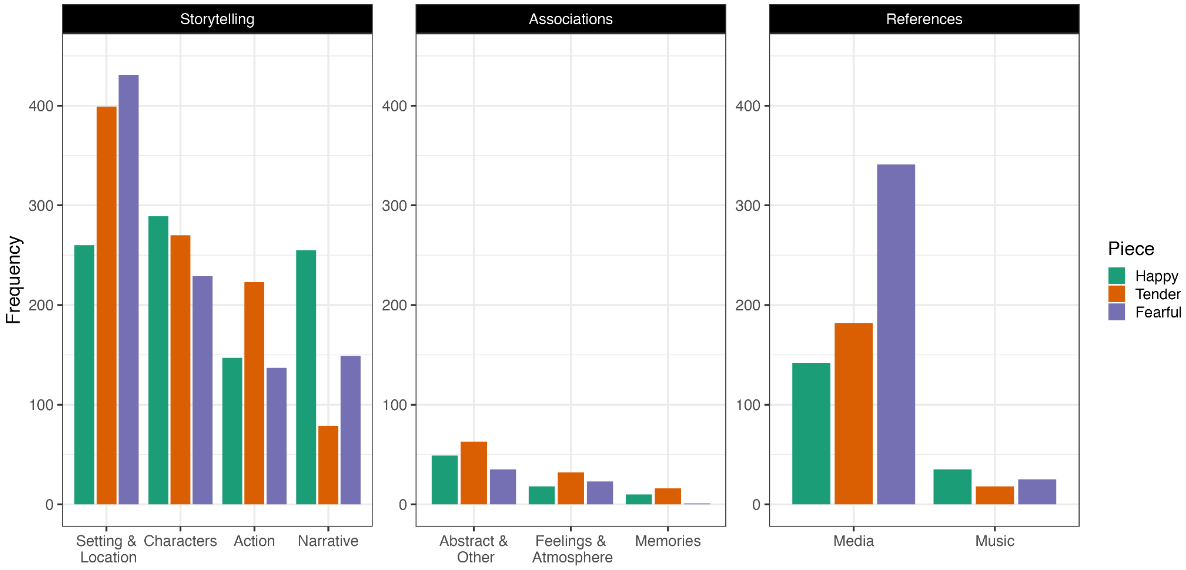


Figure A.2.

Frequencies of level 3 visual imagery content, segmented by level 2 groupings.

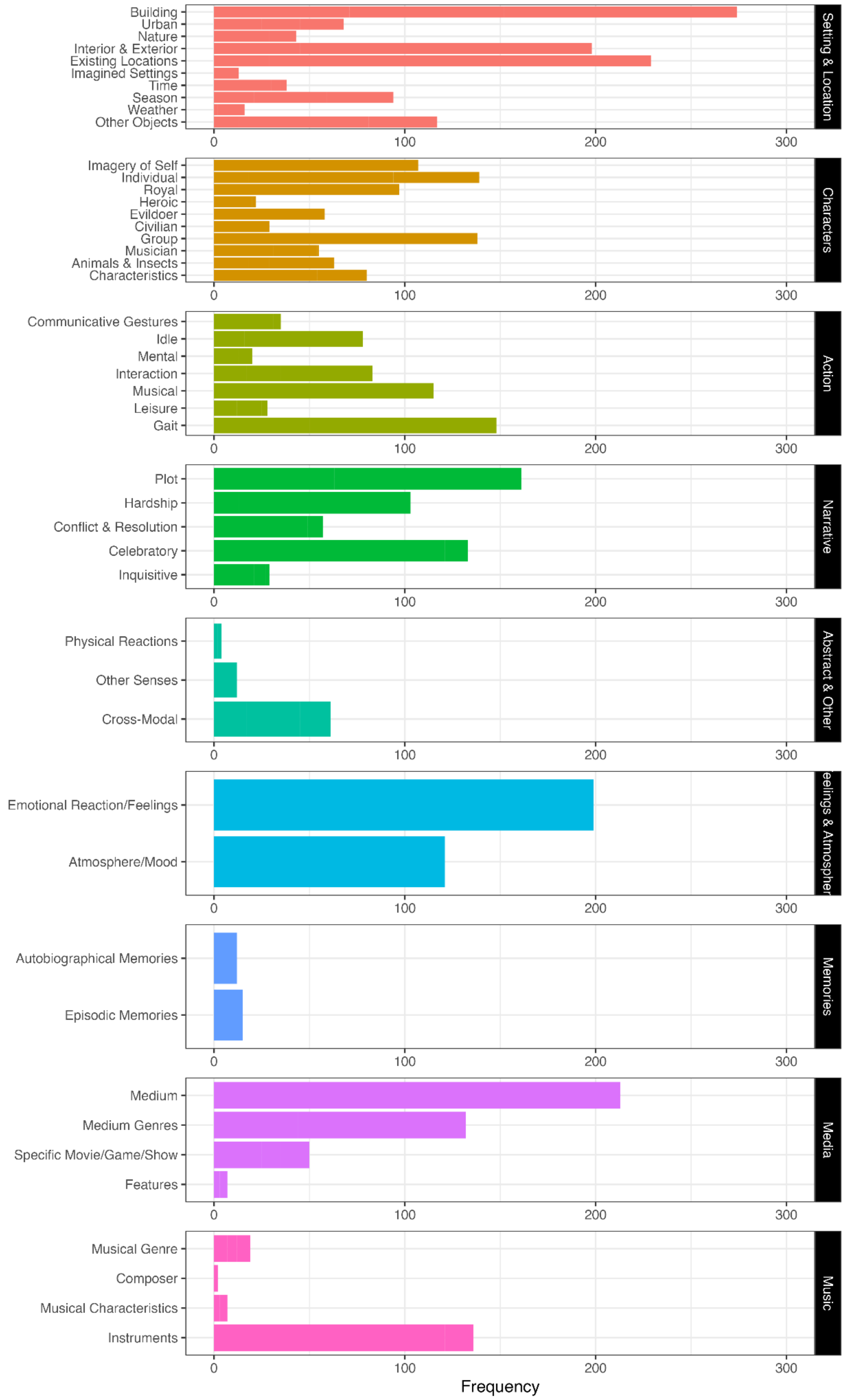


Figure A.3.

Comparing density distributions of consistency values of within- (across listening situations) and across-participant (across listeners) groups for each music excerpt type (with mean intercepts). (A) Level 2 within-participants consistency values. (B) Level 3 within-participants consistency values. (C) Level 2 across-participants consistency values. (D) Level 3 across-participants consistency values.

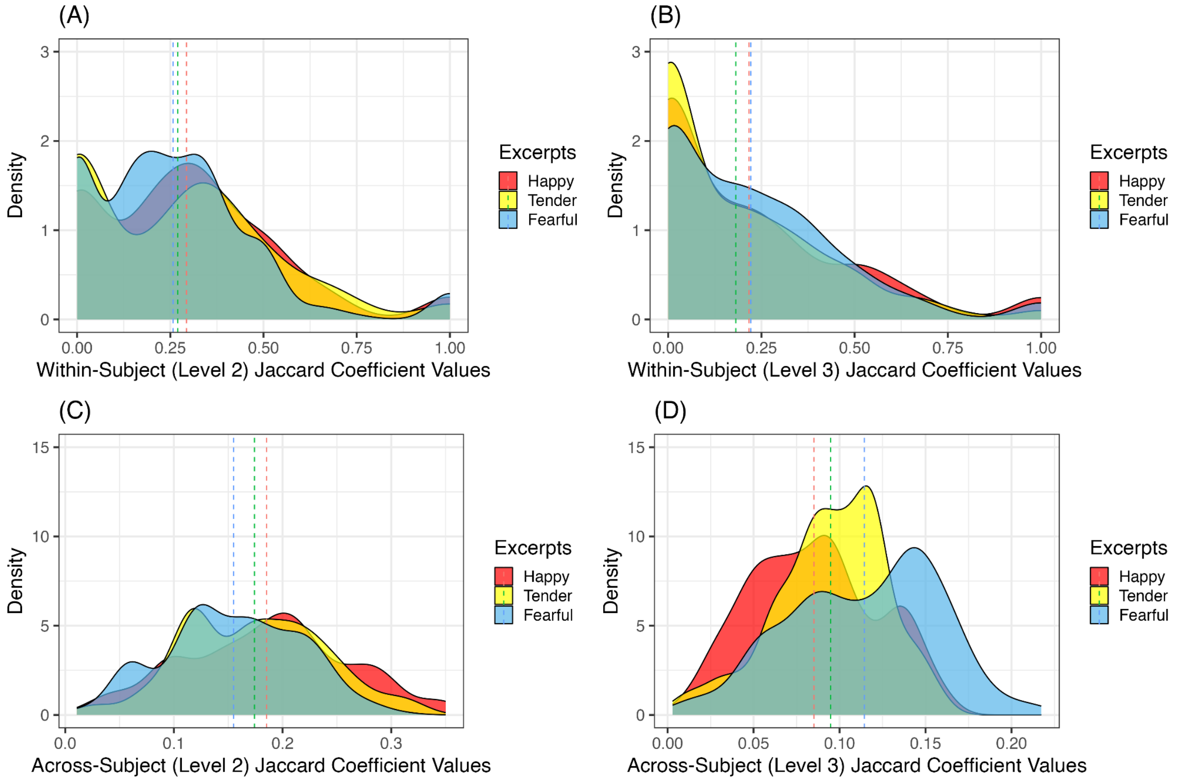


Figure A.4.

Comparing density distributions of consistency values of within- (across listening situations) and across-participant (across listeners) groups, as well as for each music excerpt type (with mean intercepts) for Storytelling codes. (A) Level 2 Storytelling codes. (B) Level 3 Storytelling codes. (C) Level 2 within-participants consistency values. (D) Level 3 within-participants consistency values. (E) Level 2 across-participants consistency values. (F) Level 3 across-participants consistency values.

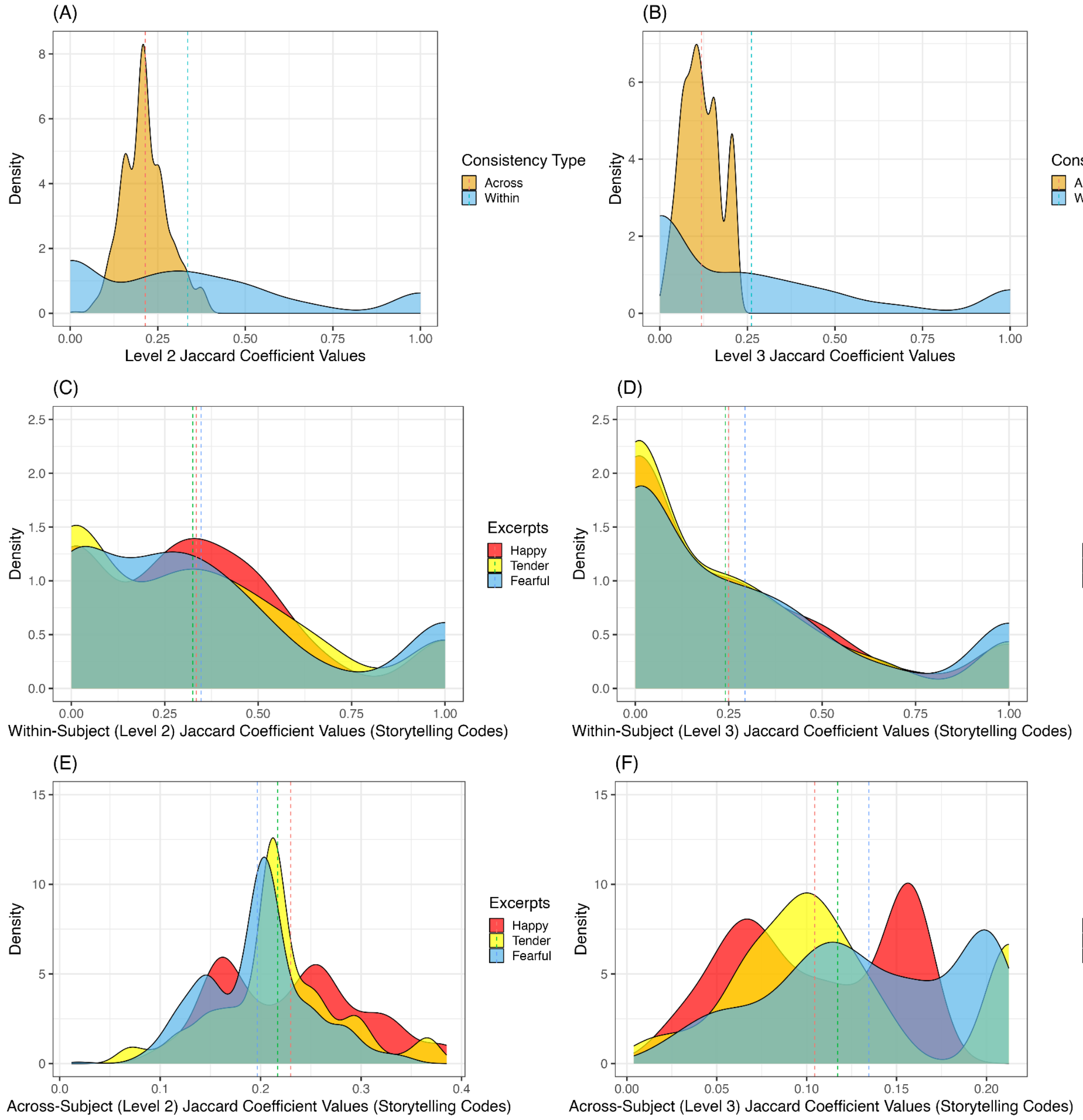


Table A.1.

Banded within-participant Jaccard consistency values across levels 2 and 3 and across music excerpts.

	Within-Participants							
	Level 3 (% of overall)				Level 2 (% of overall)			
	Overall	Happy	Tender	Fearful	Overall	Happy	Tender	Fearful
0%	300 (41.7%)	104 (43.3%)	112 (46.7%)	84 (35.0%)	195 (27.1%)	59 (24.6%)	82 (34.2%)	54 (22.5%)
0.01–19.99%	99 (13.8%)	30 (12.5%)	31 (12.9%)	38 (15.8%)	79 (11.0%)	22 (9.2%)	16 (6.7%)	41 (17.1%)
20–39.99%	178 (24.7%)	55 (22.9%)	54 (22.5%)	69 (28.7%)	255 (35.4%)	89 (37.1%)	72 (30.0%)	94 (39.2%)
40–59.99%	84 (11.7%)	26 (10.8%)	27 (11.2%)	31 (12.9%)	122 (16.9%)	44 (18.3%)	42 (17.5%)	36 (15.0%)
60–79.99%	35 (4.9%)	14 (5.8%)	11 (4.6%)	10 (4.2%)	40 (5.6%)	15 (6.2%)	19 (7.9%)	6 (2.5%)
80–99.99%	1 (0.1%)	0 (0.0%)	1 (0.4%)	0 (0.0%)	1 (0.1%)	0 (0.0%)	1 (0.4%)	0 (0.0%)
100%	23 (3.2%)	11 (4.6%)	4 (1.7%)	8 (3.3%)	28 (3.9%)	11 (4.6%)	8 (3.3%)	9 (3.8%)

Note. Within-participant total % calculated from N = 720. From total % for music excerpts were computed using each individual N sizes; $N_{\text{Happy}} = 240$, $N_{\text{Tender}} = 240$, $N_{\text{Fearful}} = 240$.

Table A.2.

Banded across-participant Jaccard consistency values across levels 2 and 3 and across music excerpts.

	Across-Participants							
	Level 3 (% of overall)				Level 2 (% of overall)			
	Overall	Happy	Tender	Fearful	Overall	Happy	Tender	Fearful
0%	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
0.01–4.99%	126 (11.9%)	71 (20.1%)	33 (9.3%)	22 (6.2%)	33 (3.1%)	19 (5.4%)	9 (2.5%)	5 (1.4%)
5–9.99%	432 (40.8%)	169 (47.9%)	148 (41.9%)	115 (32.6%)	130 (12.3%)	49 (13.9%)	24 (6.8%)	57 (16.1%)
10–14.99%	402 (38.0%)	101 (28.6%)	164 (46.5%)	137 (38.8%)	242 (22.9%)	43 (12.2%)	98 (27.8%)	101 (28.6%)
15–19.99%	94 (8.9%)	12 (3.4%)	8 (2.3%)	74 (21.0%)	276 (26.1%)	87 (24.6%)	91 (25.8%)	98 (27.8%)
20–24.99%	5 (0.5%)	-	-	5 (1.4%)	248 (23.4%)	80 (22.7%)	91 (25.8%)	77 (21.8%)
25-29.99%	-	-	-	-	99 (9.3%)	60 (17.0%)	25 (7.1%)	14 (4.0%)
30-34.99%	-	-	-	-	25 (2.4%)	9 (2.5%)	15 (4.2%)	1 (0.3%)
35-40%	-	-	-	-	6 (0.6%)	6 (1.7%)	-	-

Note. Across-participant total % were calculated from N = 1059. From total % for music excerpts were computed using each individual N sizes; N_{Happy} = 353, N_{Tender} = 353, $N_{Fearful}$ = 353.

Table A.3.

Within-participant Jaccard consistency values including only Storytelling codes across levels 2 and 3 and across excerpts.

	Within-Participants							
	Level 3 (% of overall)				Level 2 (% of overall)			
	Overall	Happy	Tender	Fearful	Overall	Happy	Tender	Fearful
0%	319 (44.3%)	110 (45.8%)	109 (45.4%)	100 (41.7%)	217 (30.1%)	67 (27.9%)	82 (34.2%)	68 (28.3%)
0.01–19.99%	61 (8.5%)	18 (7.5%)	25 (10.4%)	18 (7.5%)	42 (5.8%)	11 (4.6%)	12 (5.0%)	19 (7.9%)
20–39.99%	145 (20.1%)	50 (20.8%)	48 (20.0%)	47 (19.6%)	194 (26.9%)	71 (29.6%)	56 (23.3%)	67 (27.9%)
40–59.99%	83 (11.5%)	29 (12.1%)	24 (10.0%)	30 (12.5%)	133 (18.5%)	53 (22.1%)	41 (17.1%)	39 (16.2%)
60–79.99%	32 (4.4%)	10 (4.2%)	12 (5.0%)	10 (4.2%)	47 (6.5%)	14 (5.6%)	23 (9.6%)	10 (4.2%)
80–99.99%	1 (0.1%)	1 (0.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
100%	79 (11.0%)	22 (9.2%)	22 (9.2%)	35 (14.6%)	87 (12.1%)	24 (10.0%)	26 (10.8%)	37 (15.4%)

Note. Within-participant total % calculated from N = 720. From total % for musical excerpts were computed using each individual N sizes; $N_{\text{Happy}} = 240$, $N_{\text{Tender}} = 240$, $N_{\text{Fearful}} = 240$.

Table A.4.

Across-participant Jaccard consistency values including only Storytelling codes across levels 2 and 3 and across excerpts.

	Across-Participants							
	Level 3 (% of overall)				Level 2 (% of overall)			
	Overall	Happy	Tender	Fearful	Overall	Happy	Tender	Fearful
0%	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
0.01–4.99%	99 (9.3%)	42 (11.9%)	29 (8.2%)	28 (7.9%)	2 (0.2%)	1 (0.3%)	0 (0.0%)	1 (0.3%)
5–9.99%	322 (30.4%)	128 (36.3%)	128 (36.3%)	66 (18.7%)	21 (2.0%)	1 (0.3%)	14 (4.0%)	6 (1.7%)
10–14.99%	308 (29.1%)	82 (23.2%)	116 (32.9%)	110 (31.2%)	142 (13.4%)	23 (6.5%)	39 (11.0%)	80 (22.7%)
15–19.99%	168 (15.9%)	101 (28.6%)	5 (1.4%)	62 (17.6%)	215 (20.3%)	113 (32.0%)	46 (13.0%)	56 (15.9%)
20–24.99%	162 (15.3%)	-	75 (21.2%)	87 (24.6%)	391 (36.9%)	64 (18.1%)	169 (47.9%)	158 (44.8%)
25-29.99%	-	-	-	-	182 (17.2%)	88 (24.9%)	53 (15.0%)	41 (11.6%)
30-34.99%	-	-	-	-	73 (6.9%)	45 (12.7%)	17 (4.8%)	11 (3.1%)
35-40%	-	-	-	-	33 (3.1%)	18 (5.1%)	15 (4.2%)	-

Note. Across-participant total % were calculated from N = 1,059. From total % for musical excerpts were computed using each individual N sizes; $N_{Happy} = 353$, $N_{Tender} = 353$, $N_{Fearful} = 353$.

Table A.5.

Independent samples t-tests of prevalence of music-evoked visual imagery ratings between those who do and those who do not participate in the visual arts.

	Visual Arts		**No Visual Arts**				
	M	**SD**	**M**	**SD**	***df***	***t***	***p***
All	4.63	1.23	4.00	1.40	119.5	3.79	**< 0.001**
Happy	4.94	1.60	4.34	1.81	119.8	2.74	**0.007**
Tender	4.87	1.56	4.03	1.85	124.68	3.88	**< 0.001**
Fearful	5.13	1.43	4.58	1.71	125.01	2.74	**0.007**

Note. Bolded p-values indicate statistical significance at the cut-off 0.05.

Table A.6.

Independent samples t-tests of vividness of music-evoked visual imagery ratings between those who do and those who do not participate in the visual arts

	Visual Arts		**No Visual Arts**				
	M	**SD**	**M**	**SD**	***df***	***t***	***p***
All	4.31	1.21	3.74	1.44	124.2	3.37	**< 0.001**
Happy	4.61	1.57	4.15	1.81	121.38	2.12	**0.036**
Tender	4.37	1.54	3.88	1.93	131.36	2.24	**0.027**
Fearful	4.68	1.54	4.19	1.74	120.08	2.29	**0.023**

Note. Bolded p-values indicate statistical significance at the cut-off 0.05.

Appendix B. Chapter 3 Supplementary Tables, Figures, and Analyses

Table B.1.

English translations of the German instructions and scale anchor points used during the experiment.

English Version	German Version
As soon as the piece of music has finished, please open your eyes and rate how strongly it evoked still and moving images in your mind's eye.	*Sobald das Musikstück beendet ist, öffnen Sie bitte Ihre Augen und beurteilen Sie, wie stark es unbewegte und bewegte Bilder vor Ihrem inneren Auge hervorgerufen hat.*
Did not trigger any still images / a lot of still images in me.	*Das Musikstück hat überhaupt keine unbewegten Bilder / sehr viele unbewegte Bilder in mir ausgelöst.*
Did not trigger any moving images / a lot of moving images in me.	*Das Musikstück hat überhaupt keine bewegten Bilder / sehr viele bewegte Bilder in mir ausgelöst.*
So that you can always close your eyes in time, a visual countdown will appear before each piece of music, announcing the beginning of the respective piece of music.	*Damit Sie Ihre Augen immer rechtzeitig schließen können, erscheint vor jedem Musikstück ein visueller Countdown, der den Beginn des jeweiligen Musikstücks ankündigt.*

Table B.2.

Analyses of variance of the main models presenting oscillatory power across the five frequency bands (alpha, including lower and upper alpha, beta, and gamma) associated with static and dynamic imagery, as well as their interactions with Time Period (TP: First Half and Second Half) and Region of Interest (ROI: Frontal, Centro-Parietal, and Parieto-Occipital).

	Alpha (8-13 Hz)				**Lower Alpha (8-10 Hz)**				**Upper Alpha (11-13 Hz)**				**Beta (14-30 Hz)**				**Gamma (30-45 Hz)**			
	F	*DF*	*DenDF*	*p*	*F*	*DF*	*DenDF*	*p*	*F*	*DF*	*DenDF*	*p*	*F*	*DF*	*DenDF*	*p*	*F*	*DF*	*DenDF*	*p*
Static Imagery	4.89	1	435460	**0.027***	4.99	1	441923	**0.025***	1.54	1	428599	**0.214**	176.64	1	453488	**< 0.001*****	0.51	1	451494	**0.475**
Dynamic Imagery	21.79	1	728	**< 0.001*****	17.53	1	6733	**< 0.001*****	11.89	1	542	**< 0.001*****	9.50	1	281826	**0.002****	5.73	1	14847	**0.017***
Time Period	151.35	1	453518	**< 0.001*****	5.48	1	453518	**0.019***	543.15	1	453518	**< 0.001*****	9.92	1	453518	**0.002****	0.81	1	453518	**0.367**
ROI	295.10	2	453518	**< 0.001*****	455.19	2	453518	**< 0.001*****	10.22	2	453518	**< 0.001*****	22.69	2	453518	**< 0.001*****	475.81	2	453518	**< 0.001*****
Static * TP	32.62	1	453518	**< 0.001*****	14.55	1	453518	**< 0.001*****	40.78	1	453518	**< 0.001*****	90.52	1	453518	**< 0.001*****	1.23	1	453518	**0.268**
Dynamic * TP	38.28	1	453518	**< 0.001*****	26.51	1	453518	**< 0.001*****	27.73	1	453518	**< 0.001*****	99.62	1	453518	**< 0.001*****	14.89	1	453518	**< 0.001*****
Static * ROI	33.09	2	453518	**< 0.001*****	10.40	2	453518	**< 0.001*****	58.92	2	453518	**< 0.001*****	6.67	2	453518	**0.001****	222.20	2	453518	**< 0.001*****
Dynamic * ROI	121.88	2	453518	**< 0.001*****	63.76	2	453518	**< 0.001*****	142.38	2	453518	**< 0.001*****	7.17	2	453518	**< 0.001*****	24.45	2	453518	**< 0.001*****
Time Period * ROI	8.01	2	453518	**< 0.001*****	0.40	2	453518	**0.670**	35.83	2	453518	**< 0.001*****	3.59	2	453518	**0.027**	0.25	2	453518	**0.782**
Static * TP * ROI	0.20	2	453518	**0.817**	0.44	2	453518	**0.647**	0.08	2	453518	**0.926**	1.06	2	453518	**0.346**	1.37	2	453518	**0.253**
Dynamic * TP * ROI	2.33	2	453518	**0.098**	1.97	2	453518	**0.139**	3.10	2	453518	**0.044***	9.19	2	453518	**< 0.001*****	0.68	2	453518	**0.508**

Note. Abbreviations: TP = Time Period, ROI = Region of Interest, DF = Degrees of Freedom, DenDf = Denominator Degrees of Freedom. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table B.3.

Summarising the descriptive statistics of lower and upper alpha power across the two time periods and the three regions of interest.

	Mean (Standard Deviation)				
	Time Period		**ROI**		
	First Half	Second Half	Frontal	Centro-Parietal	Parieto-Occipital
Lower Alpha	–5.60 (51.47)	–5.22 (51.75)	–7.38 (49.26)	–4.43 (52.36)	–4.41 (53.26)
Upper Alpha	–2.55 (17.71)	–3.82 (16.32)	–2.91 (13.28)	–2.75 (17.71)	–4.19 (19.92)

Figure B.1.

Frequency power plot displaying oscillatory power across all trials. Individual colour-coded lines represent each of the EEG channels.

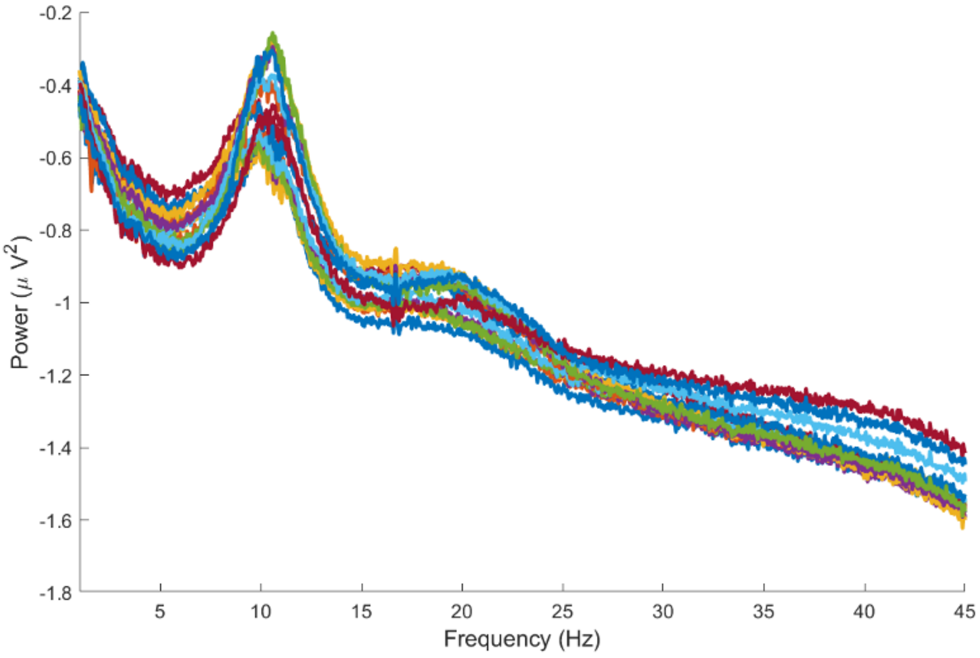
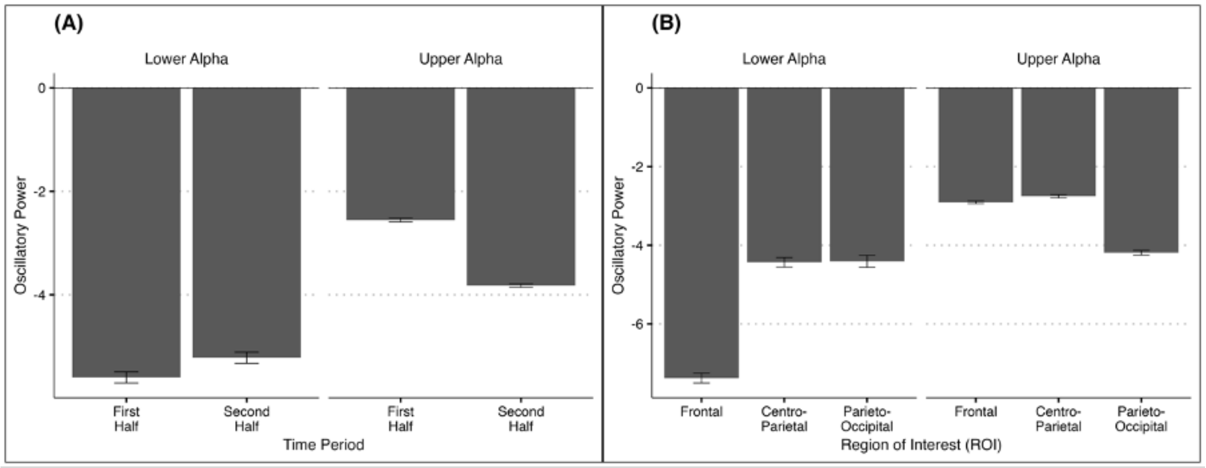


Figure B.2.

Illustrating oscillatory power of lower and upper alpha, including ± standard error of the mean. (A) Depicting average power across the two time periods (First Half and Second Half). (B) Depicting average power across the three regions of interest (ROI; Frontal, Centro-Parietal, and Parieto-Occipital).



Exploratory analyses: lower and upper alpha (*continued*)

The overall model for lower alpha revealed a main effect of static, F(1, 441,923) = 4.99, p = 0.025, as well as dynamic imagery, F(1, 6,733) = 17.53, p < 0.001, whereby high ratings were associated with suppression in the low alpha frequencies. There was a further main effect of Time Period, F(2, 453,518) = 5.48, p = 0.019, whereby lower alpha was slightly more suppressed in the first half than in the second, and of ROI, F(2, 453,518) = 455.19, p < 0.001, whereby the frontal area showed most suppression in the low alpha frequencies, followed by the centro-parietal area, then the parieto-occipital area. There were also significant interactions between static imagery and time period, F(1, 453,518) = 14.55, p < 0.001, and between dynamic imagery and time period, F(1, 453,518) = 26.51, p < 0.001. To explore these interactions, models were run for static and dynamic imagery and each time period separately. Static imagery showed a significant negative relationship with lower alpha power in the first half, ß = –0.00795, SE = 0.00192, t(2.14) = –4.13, p < 0.001, and a significant positive relationship in the second half, ß = 0.00492, SE = 0.00194, t(1.67) = 2.53, p = 0.011. Further, there were significant negative relationships between dynamic imagery and lower alpha in the first half, ß = –0.00626, SE = 0.00185, t(1.27) = –3.40, p < 0.001, but a non-significant effect in the second half, ß = –0.00218, SE = 0.00185, t(633.71) = –1.18, p = 0.238.

I further found significant interactions between static imagery and ROI, F(2, 453,518) = 10.40, p < 0.001, and between dynamic imagery and ROI, F(2, 453,518) = 63.76, p < 0.001. Follow up models were run for static and dynamic imagery and each ROI separately. With regard to ROI, there were non-significant negative relationships between static imagery and the frontal, ß = 0.00019, SE = 0.00124, t(1.18) = –0.15, p = 0.877, centro-parietal, ß = –0.00147, SE = 0.00240, t(1.57) = –0.61, p = 0.541, and parieto-occipital areas, ß = –0.00340, SE = 0.00296, t(1.00) = –1.15, p = 0.251. In addition, there was a significant negative relationship between dynamic imagery and power in the frontal area, ß = –0.00313, SE = 0.00119, *t*(1568.03) = –2.63, *p* = 0.009, but non-significant negative relationships in the centro-parietal, ß = –0.00402, SE = 0.00223, *t*(2.31) = –1.80, *p* = 0.072, and parieto-occipital areas, ß = –0.00344, SE = 0.00264, *t*(76.47) = –1.30, *p* = 0.196. No other main effects or interactions were significant.

In the main model assessing upper alpha power, there was a significant main effect of dynamic imagery, *F*(1, 542) = 11.89, *p* < 0.001, whereby high ratings were associated with suppressed upper alpha power. There was also a significant main effect of Time Period, *F*(1, 453,518) = 543.16, *p* < 0.001, whereby there was less upper alpha in the second compared to the first half of the trial, as well as a main effect of ROI, *F*(2, 453,518) = 10.22, *p* < 0.001, whereby power was most suppressed in the parieto-occipital area than in the frontal and centro-parietal areas. There were significant interactions found between static imagery and time period, *F*(1, 453,518) = 40.78, *p* < 0.001, and between dynamic imagery and time period, *F*(1, 453,518) = 27.73, *p* < 0.001. Models were run for static and dynamic imagery and each time period separately to explore these interactions further. There was a non-significant negative relationship between static imagery and the first half, ß = 0.00088, SE = 0.00128, *t*(1.92) = –0.69, *p* = 0.488, and a non-significant positive relationship with the second half, ß = 0.00164, SE = 0.00110, *t*(2.09) = 1.50, *p* = 0.134. Further, there was a significant negative relationship between dynamic imagery and the first half, ß = –0.00457, SE = 0.00119, *t*(2.67) = –3.84, *p* < 0.001, and a non-significant positive relationship between dynamic imagery and the second half, ß = 0.00147, SE = 0.00106, *t*(3.67) = 1.39, *p* = 0.163.

There were further significant interactions between static imagery and ROI, *F*(2, 453,518) = 58.92, *p* < 0.001, as well as between dynamic imagery and ROI, *F*(2, 453,518) = 142.38, *p* < 0.001. Models were run for static and dynamic imagery and each ROI separately. With regard to ROI, there was a significant negative relationship between static imagery and upper alpha power in the frontal area, ß = –0.00177, SE = 0.00066, $t(1.36) = -2.68$, $p = 0.007$, and non-significant positive relationships with the centro-parietal, ß = 0.00114, SE = 0.00142, $t(1.81) = 0.80$, $p = 0.424$, and the parieto-occipital areas, ß = 0.00155, SE = 0.00208, $t(1.06) = 0.75$, $p = 0.457$. Finally, dynamic imagery showed a significant negative relationship with the frontal area, ß = –0.00184, SE = 0.00058, $t(54.95) = -3.17$, $p = 0.003$, a non-significant positive relationship with the centro-parietal area, ß = –0.00013, SE = 0.00127, $t(6.55) = 0.10$, $p = 0.919$, and a non-significant negative relationship with the parieto-occipital area, ß = –0.00035, SE = 0.00199, $t(1.63) = -1.77$, $p = 0.077$.

There was a significant three-way interaction between dynamic imagery, time period, and ROI, $F(2, 453{,}518) = 3.10$, $p = 0.045$. Exploration of this three-way interaction revealed a significant negative relationship between dynamic imagery and upper alpha power in the frontal area in the first half, ß = –0.00367, SE = 0.00095, $t(2.27) = -3.93$, $p < 0.001$, but a non-significant negative relationship in the second half, ß = –0.00044, SE = 0.00766, $t(1.26) = -0.58$, $p = 0.565$.

Further, there was a non-significant negative relationship between dynamic imagery and upper alpha power in the centro-parietal area in the first half, ß = –0.00237, SE = 0.00201, $t(350.44) = -1.18$, $p = 0.240$, and a non-significant positive relationship in the second half, ß = 0.00262, SE = 0.00179, $t(593.50) = 1.46$, $p = 0.144$. There was also a significant negative relationship between dynamic imagery and upper alpha power in the parieto-occipital area in the first half, ß = –0.00877, SE = 0.00298, $t(231.20) = -2.95$, $p = 0.004$, as well as a non-significant positive relationship in the second half, ß = 0.00255, SE = 0.00248, $t(1.49) = 1.03$, $p = 0.306$.

Finally, there was a significant interaction between ROI and time period, $F(2, 453{,}518) = 35.84$, $p < 0.001$. However, this was not explored further due to the current investigation's focus towards observing visual imagery effects. No other main effects or interactions were significant.

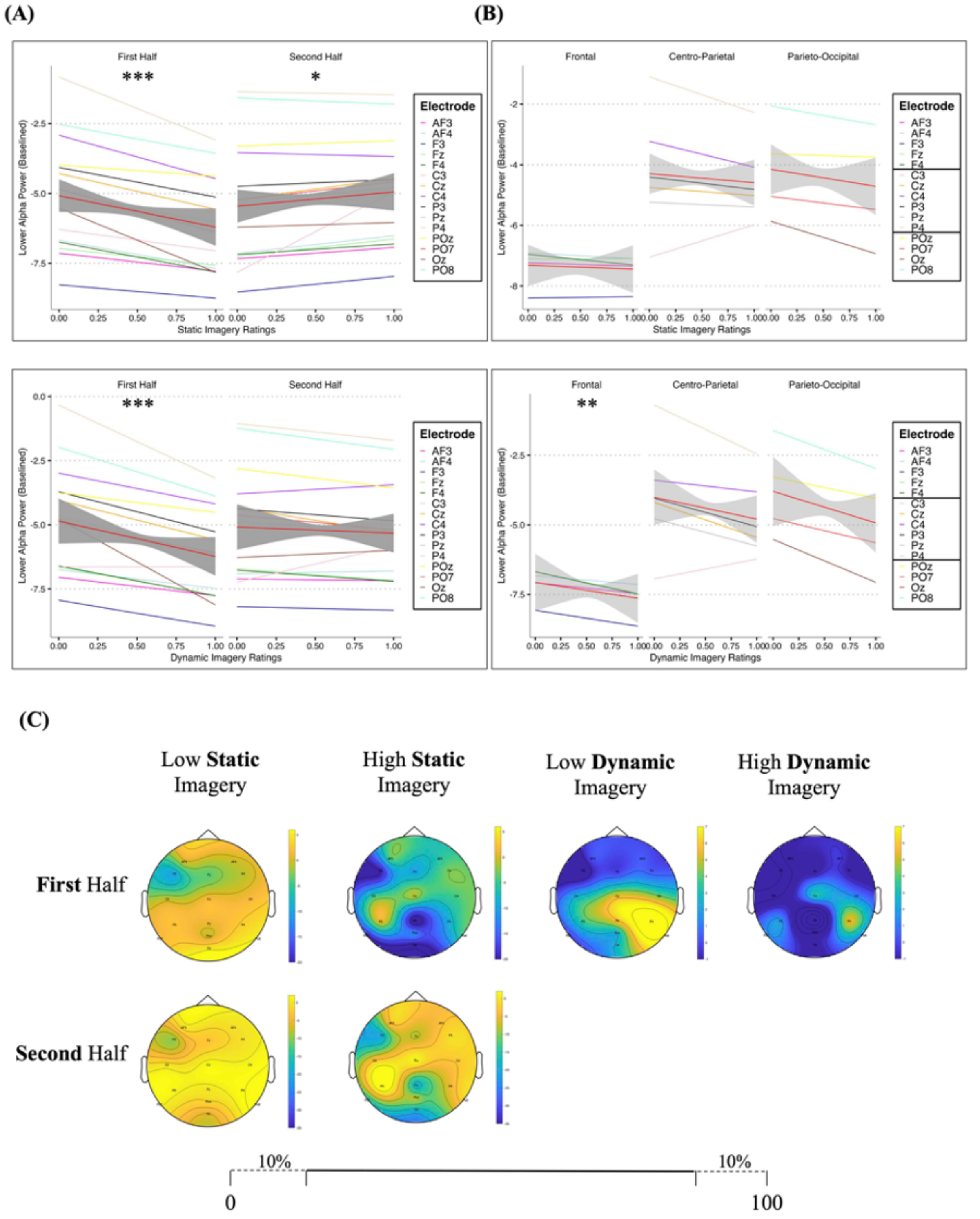


Figure B.3.

*Lower alpha power associated with static and dynamic imagery. Asterisks denote model significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (A) Scatterplots showing lower alpha power as a function of static and dynamic imagery ratings, once the other had been regressed out, for two different time windows (First Half and Second Half). For illustrative* *purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (B) Scatterplots showing lower alpha power as a function of imagery ratings, once the other rating had been regressed out, for three regions of interest (Frontal: AF3, AF4, F3, Fz, and F4; Centro-Parietal: C3, Cz, C4, P3, Pz, and P4; and Parieto-Occipital: POz, PO7, Oz, and PO8). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (C) For illustrative purposes, I present topoplots showing patterns of lower alpha power (baselined to 2 seconds prior to the music) from trials associated with the upper 10% (denoted High) and lower 10% (denoted Low) of static imagery ratings across the two different time windows, and dynamic imagery ratings for the first half only.*

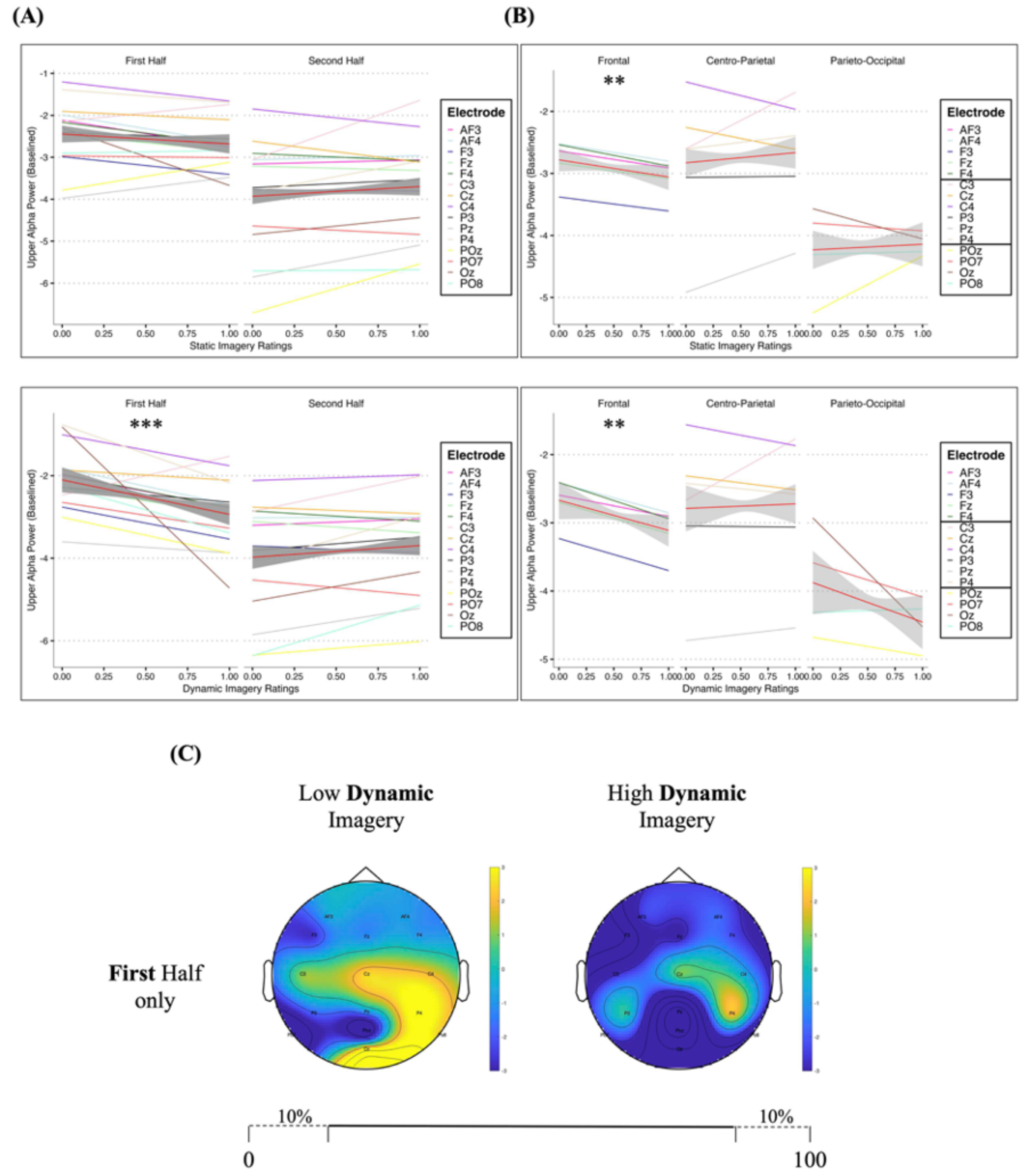


Figure B.4.

*Upper alpha power associated with static and dynamic imagery. Asterisks denote model significance levels: ** p < 0.01, *** p < 0.001. (A) Scatterplots showing upper alpha power as a function of static and dynamic imagery ratings, once the other had been regressed out, for two different time windows (First Half and Second Half). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence* *interval. (B) Scatterplots showing upper alpha power as a function of imagery ratings, once the other rating had been regressed out, for three regions of interest (Frontal: AF3, AF4, F3, Fz, and F4; Centro-Parietal: C3, Cz, C4, P3, Pz, and P4; and Parieto-Occipital: POz, PO7, Oz, and PO8). For illustrative purposes, the x-axis is scaled to between 0 and 1. Individual electrodes are represented by different colours. Thick red lines illustrate aggregated mean electrode power values, with shaded 95% confidence interval. (C) For illustrative purposes, I present topoplots showing patterns of upper alpha power (baselined to 2 seconds prior to the music) from trials associated with the upper 10% (denoted High) and lower 10% (denoted Low) of dynamic imagery ratings for the first half only.*

Exploratory analyses: effects of musicianship (*continued*)

Alpha power

The model predicting alpha power demonstrated a significant main effect of static imagery, *F*(1, 425,676) = 4.94, $p = 0.026$, dynamic imagery, *F*(1, 1,059) = 15.54, $p < 0.001$, and time period, *F*(1, 453,525) = 123.19, $p < 0.001$. There was a significant interaction found between time period and static imagery, *F*(1, 453,525) = 23.47, $p < 0.001$, as well as dynamic imagery, *F*(1, 453,525) = 32.15, $p < 0.001$. Further, there was a significant interaction between static imagery and musical training, *F*(1, 282,270) = 19.56, $p < 0.001$, and between dynamic imagery and musical training, *F*(1, 453,525) = 32.15, $p < 0.001$, as well as between time period and musical training, *F*(1, 453,525) = 37.91, $p < 0.001$. Finally, there were significant three-way interactions between musical training, time period, and dynamic imagery, *F*(1, 453,525) = 7.64, $p = 0.006$, as well as static imagery, *F*(1, 453,525) = 73.21, $p < 0.001$. No other main effects or interactions were significant. To address all possible dimensions of the effects, only the three-way interactions were analysed further with follow-up linear mixed models exploring the interaction between imagery across high and low musical training, and across the two time points separately.

In the follow-up model predicting alpha power in the first half for those with high training, there was a significant negative relationship between power and static imagery, ß = –0.00895, SE = 0.00171, *t*(9.81) = –5.25, $p < 0.001$, whereas for those with low training in the first half there was a non-significant negative relationship, ß = –0.00100, SE = 0.00207, *t*(7.87) = –0.49, $p = 0.628$. The model of the second half with those with high training showed a non-significant negative relationship between alpha power and static imagery, ß = –0.00184, SE =

0.00155, $t(1.09) = -1.18$, $p = 0.237$, whereas the model with low training revealed a significant positive relationship, ß = 0.00618, SE = 0.00200, $t(9.31) = 3.09$, $p = 0.002$.

Regarding dynamic imagery, the follow-up model predicting alpha power in the first half for those with high training showed there to be a non-significant negative relationship, ß = –0.00220, SE = 0.00149, $t(1.51) = -1.48$, $p = 0.141$, whereas for those with low training in the first half there was a significant negative relationship, ß = –0.00798, SE = 0.00207, $t(737.63) = -3.85$, $p < 0.001$. The model of the second half with those with high training showed a significant negative relationship between alpha power and dynamic imagery, ß = –0.00462, SE = 0.00137, $t(2.96) = -3.38$, $p < 0.001$, whereas the model with low training revealed a non-significant positive relationship, ß = 0.00354, SE = 0.00207, $t(3.63) = 1.71$, $p = 0.087$.

Beta power

The model predicting beta power showed a significant main effect of static imagery, $F(1, 453{,}432) = 136.72$, $p < 0.001$, dynamic imagery, $F(1, 271{,}741) = 15.57$, $p < 0.001$, and time period, $F(1, 453{,}525) = 10.09$, $p = 0.001$. There was a significant interaction between static imagery and musical training, $F(1, 452{,}889) = 61.57$, $p < 0.001$, and between static imagery and time period, $F(1, 453{,}525) = 81.39$, $p < 0.001$. Likewise, there was a significant interaction between dynamic imagery and musical training, $F(1, 453{,}542) = 55.41$, $p < 0.001$, and between dynamic imagery and time period, $F(1, 453{,}525) = 96.95$, $p < 0.001$. Further, the model revealed a three-way interaction between musical training, time period, and static, $F(1, 453{,}525) = 32.39$, $p < 0.001$, as well as dynamic imagery, $F(1, 453{,}525) = 8.34$, $p = 0.004$. No other main effects or interactions were significant.

In the follow-up model predicting beta power in the first half for those with high musical training, there was a significant positive relationship between power and static imagery, ß = 0.00444, SE = 0.00118, $t(2.37) = 3.77$, $p < 0.001$, and for those with low training in the first half, there was also a significant positive relationship, ß = 0.01741, SE = 0.00067, $t(1.13) = 26.14$, $p < 0.001$. The model of the second half with those with high training showed a non-significant negative relationship between beta power and static imagery, ß = –0.00060, SE = 0.00085, $t(4.60) = -0.70$, $p = 0.484$, whereas the model with low training revealed a significant positive relationship, ß = 0.00307, SE = 0.00083, $t(1.13) = 3.71$, $p < 0.001$.

With regard to dynamic imagery, the follow-up model predicting beta power in the first half for those with high training showed there to be a non-significant negative relationship, ß = –0.00025, SE = 0.00103, $t(7.04) = -0.24$, $p = 0.807$, whereas for those with low training in the first half there was a significant negative relationship, ß = –0.01331, SE = 0.00070, $t(8.89) = -19.08$, $p < 0.001$. The model of the second half with those with high training showed a significant positive relationship between beta power and dynamic imagery, ß = 0.00321, SE = 0.00076, $t(1.93) = 4.21$, $p < 0.001$, whereas the model with low training revealed a non-significant positive relationship, ß = 0.00135, SE = 0.00087, $t(1.01) = 1.56$, $p = 0.120$.

Gamma power

The model predicting gamma power showed there to be a significant main effect of static imagery, $F(1, 450{,}445) = 7.72$, $p = 0.005$, and dynamic imagery, $F(1, 12{,}888) = 6.08$, $p = 0.014$. There was also a significant interaction between musical training and static imagery, $F(1, 431{,}027) = 24.41$, $p < 0.001$, a significant interaction between musical training and time period, $F(1, 453{,}525) = 8.82$, $p = 0.002$, and a significant interaction between dynamic imagery and time period, $F(1, 453{,}525) = 14.71$, $p < 0.001$.

Further, there was a significant three-way interaction between musical training, time period, and static, $F(1, 453{,}525) = 4.78$, $p = 0.029$, as well as dynamic imagery, $F(1, 453{,}525) = 16.06$, $p < 0.001$. No other main effects or interactions were significant.

In the follow-up model predicting gamma power in the first half for those with high musical training, there was a significant positive relationship between power and static imagery, ß = 0.00082, SE = 0.00014, $t(1.09) = 5.91$, $p < 0.001$, and for those with low training in the first half, there was a significant negative relationship, ß = –0.00045, SE = 0.00019, $t(8.81) = –2.38$, $p = 0.018$. The model of the second half with those with high training showed a significant positive relationship between gamma power and static imagery, ß = 0.00068, SE = 0.00013, $t(1.08) = 5.15$, $p < 0.001$, whereas the model with low training revealed a non-significant positive relationship, ß = 0.00011, SE = 0.00029, $t(9.46) = 0.39$, $p = 0.697$.

With regard to dynamic imagery, the follow-up model predicting gamma power in the first half for those with high training showed there to be a non-significant positive relationship, ß = 0.00004, SE = 0.00012, $t(4.72) = 0.34$, $p = 0.736$, whereas for those with low training in the first half there was a significant positive relationship, ß = 0.00070, SE = 0.00019, $t(2.56) = 3.63$, $p < 0.001$. The model of the second half with those with high training showed a non-significant positive relationship between gamma power and dynamic imagery, ß = 0.00006, SE = 0.00011, $t(4.29) = 0.49$, $p = 0.621$, whereas the model with low training revealed a non-significant negative relationship, ß = –0.00011, SE = 0.00030, $t(2.87) = –0.36$, $p = 0.717$.

Appendix C. Chapters 4 and 5 Pilot Data, and Supplementary Tables and Figures

Pilot survey for the selection of unfamiliar music of high and low relaxation potential (*continued)*

I ran paired samples t-tests to examine the differences between the tracks included in the pilot study along a set of acoustic qualities (energy, acousticness, valence, danceability, and loudness) as well as a couple of aesthetic qualities (relaxingness and liking) and familiarity, for the purpose of selecting a final set of tracks to be presented as part of the main study's unfamiliar relaxing and non-relaxing listening conditions. The intention was to select one track from within each genre (EDM, classical, and jazz) from within the low relaxation potential (non-relaxing listening condition) and high relaxation potential (relaxing listening condition) categories. To this end, each of the two tracks within the same genre under each relaxation potential category were compared, according to the criteria outlined in section *4.2.4*. Bonferroni correction was applied to account for the seven comparisons made within each genre type in the sets of high and low relaxation potential tracks, resulting in an adjusted alpha of 0.007. See Table C.2 below for the means and standard deviations of the quality and familiarity ratings for each track.

With regard to the high relaxation potential tracks, paired samples t-tests showed there to be a significant difference between the electronic dance music (EDM) tracks *Escape (Fictivision Mix)* and *Untitled Instrumental No.4* in terms of valence only, $t(16) = 3.12$, $p = 0.007$, with the *Escape (Fictivision Mix)* track being rated with more positive valence. The classical music tracks *Carnival Overture (Op. 92)* and *Piano Concerto No. 1 in E-flat Major* showed significant differences in terms of energy, $t(16) = 3.52$, $p = 0.003$, and valence, $t(16) =$

3.12, $p = 0.007$, with the *Carnival Overture* track performing higher in both indices. Further, the jazz music tracks showed significant differences in terms of acousticness, $t(16) = 3.23$, $p = 0.005$, and valence, $t(16) = 3.10$, $p = 0.007$, the track *Flik's Machine* being rated with more positive valence and as being more acoustic than the *Quest for Coin* track.

Next, in terms of the low relaxation potential tracks, paired samples t-tests revealed significant differences between the EDM tracks *Ambient Track* and *Akiko* in energy, $t(16) = 6.43$, $p < 0.001$, valence, $t(16) = 3.82$, $p = 0.002$, and danceability, $t(16) = 7.35$, $p < 0.001$, with *Ambient Track* being considered as possessing low energy, neutral valence, and low danceability. Further, paired samples t-tests between the classical music tracks *Christmas Oratorio, BWV 248* and *Berlioz: Symphonie Fantastique* did not reveal any significant differences in terms of any of the track quality indices, however showed a tendency for the *Christmas Oratorio* track to be considered less energetic, neutrally valenced, more relaxing, and was liked more, which fulfilled our selection criteria to a greater extent than the *Symphonie Fantastique*. Finally, there was a significant difference between the jazz tracks *O Leazinho* and *Hello My Lovely* in terms of valence only, $t(16) = 3.36$, $p = 0.004$, with *O Leazinho* being rated as possessing more positive valence.

In a final check, I ran three additional paired-samples t-tests, this time comparing relaxingness ratings *across* conditions, in order to be sure that our chosen tracks within each genre indeed differed in terms of their relaxation potential. I found a significant differences in levels of relaxingness between our EDM tracks, *Escape (Fictivision Mix)* and *Ambient Track*, $t(16) = 7.17$, $p < 0.001$, between our classical tracks, *Carnival Overture (Op. 92)* and *Berlioz: Symphonie Fantastique*, $t(16) = 4.00$, $p = 0.001$, as well as between our jazz tracks, *Flik's Machine* and *Hello My Lovely*, $t(16) = 3.87$, $p = 0.001$.

Table C.1.

Musical excerpts included in the pilot survey with identifying details.

Piece	Artist/Composer	Genre	Arousal
Escape (Fictivision Mix)*^	Fictivision	Electronic	High
Untitled Instrumental No.4	Antoni Maiovvi	Electronic	High
Carnival Overture (Op. 92)*^	Antonín Dvořák	Classical	High
Piano Concerto No. 1 in E-flat Major, S. 124 III Allegretto Vivace^	Franz Liszt	Classical	High
The Flik Machine*^	Randy Newman	Jazz	High
Quest for Coin	Ezra Collective	Jazz	High
Ambient Track*	MassiveMusic	Electronic	Low
Akiko^	Guitar	Electronic	Low
Christmas Oratorio, BWV 248: Sinfonia in G*^	Various Artists	Classical	Low
Berlioz: Symphonie Fantastique, Op. 14 – 3. Scene Aux Champs*^	Michael Tilson Thomas: San Francisco Symphony Orchestra	Classical	Low
O Leazinho^	Brooks Williams	Jazz	Low
Hello My Lovely*^	Enrico Pieranunzi	Jazz	Low

Note. Those marked with * were the final tracks chosen for the main experiment. In terms of tracks selected to be piloted, those marked with ^ were selected from Marti-Marca et al. (2020), and unmarked excerpts were selected by the experimenter independently.

Table C.2.

Mean values (and standard deviations) of acoustic and aesthetic ratings of musical excerpts compared in a pilot survey for the purpose of selecting tracks to be used in the experimenter-selected conditions for the main experiment.

	Genre	Energy	Acousticness	Valence	Danceability	Loudness	Relaxingness	Liking	Familiarity
High Relaxation Potential									
Escape (Fictivision Mix) *	Electronic	6.76 (0.44)	1.47 (1.46)	5.76 (1.30)	6.00 (1.84)	6.06 (1.03)	1.59 (1.06)	4.29 (2.08)	1.06 (0.24)
Antoni Maiovvi - Untitled Instrumental No.4	Electronic	6.65 (0.61)	1.35 (1.06)	4.88 (0.86)	5.35 (1.87)	5.88 (1.17)	1.53 (1.07)	3.71 (1.83)	1.12 (0.33)
Dvorak, Carnival Overture *	Classical	6.29 (1.16)	5.76 (2.08)	5.29 (1.86)	2.76 (2.19)	5.82 (1.29)	2.12 (1.45)	4.00 (1.87)	1.76 (0.90)
Liszt, Allegretto Vivace	Classical	5.29 (1.21)	5.65 (1.84)	3.65 (1.90)	1.88 (1.53)	5.59 (1.54)	1.94 (1.34)	4.18 (1.94)	1.53 (0.80)
Flik's Machine *	Jazz	6.47 (0.72)	5.76 (1.92)	6.76 (0.44)	5.82 (1.47)	5.47 (1.07)	3.35 (1.80)	5.65 (1.54)	1.41 (0.62)
Quest for Coin	Jazz	6.12 (1.11)	4.47 (1.70)	6.12 (0.93)	5.71 (1.76)	5.41 (1.42)	3.24 (1.44)	5.24 (1.99)	1.06 (0.24)
Low Relaxation Potential									
Ambient Track *	Electronic	1.64 (1.17)	3.47 (2.00)	4.18 (1.78)	1.24 (0.56)	2.71 (1.49)	5.47 (2.07)	3.76 (1.95)	1.12 (0.33)
Akiko	Electronic	4.24 (1.03)	3.82 (1.59)	5.24 (1.09)	3.76 (1.64)	3.53 (1.12)	4.82 (1.98)	4.94 (1.89)	1.12 (0.33)
Christmas Oratorio, BWV 248	Classical	3.06 (1.14)	6.12 (1.53)	4.88 (1.45)	3.06 (1.89)	4.12 (1.76)	5.24 (1.89)	4.47 (1.81)	1.47 (0.72)
Berlioz: Symphonie Fantastique *	Classical	3.41 (1.77)	5.88 (1.58)	4.65 (1.41)	2.41 (1.73)	4.18 (1.38)	4.41 (2.03)	4.00 (2.03)	1.47 (0.51)
O Leazinho	Jazz	4.06 (1.03)	6.47 (0.94)	5.59 (1.12)	2.18 (1.07)	2.94 (1.34)	5.53 (1.84)	5.00 (1.65)	1.12 (0.33)
Hello My Lovely *	Jazz	2.65 (1.46)	5.29 (2.23)	4.06 (1.48)	2.59 (1.50)	2.88 (1.45)	5.41 (1.97)	4.24 (1.92)	1.18 (0.53)

Note. N = 17. Excerpt pairs within each genre were compared using paired-samples t-tests for inclusion in the final set of unfamiliar tracks. Those marked with * were the final tracks chosen for the main experiment.

Table C.3.

Additional questions shown in Studies 3 and 4 regarding visual imagery and overall listening experience.

Post-Listening Questions
On average...
...how much did you like the pieces of music (track) you just heard?*
...how relaxing did you find the pieces of music (track) you just heard?*
...were the contents of your thoughts during the music positive or negative?
*...how anxious did the sounds and numbers task make you?**
*...how much time during the listening did you spend thinking about the sounds and numbers task?**
Indicate the extent to which any imagery...
...followed a narrative or story-line
...involved characters or people you don't know
...involved an imagined location or setting (i.e., indoors, outdoors)
...was based on past memories or events
...were thoughts about yourself
...included people you know (e.g., family, friends, partner)
...included abstract content (e.g., shapes, colours)
...involved yourself or others performing any actions or movements
Indicate the extent to which you were...
...thinking about the music (e.g., its melody, beat, harmony)
...thinking about the experiment

Note. Those marked with * were questions only included or phrased differently in Study 4. Remaining unmarked items were included in both Studies 3 and 4. All items were rated on 1-7 Likert scale.

Figure C.1.

Mean values (± standard error of the mean) of track quality responses for high relaxation potential tracks compared within genres (top plot: electronic dance music genre tracks, middle plot: classical music genre tracks, bottom plot: jazz music genre tracks). Asterisks indicate differences that reached corrected statistical significance ($p < 0.007$).

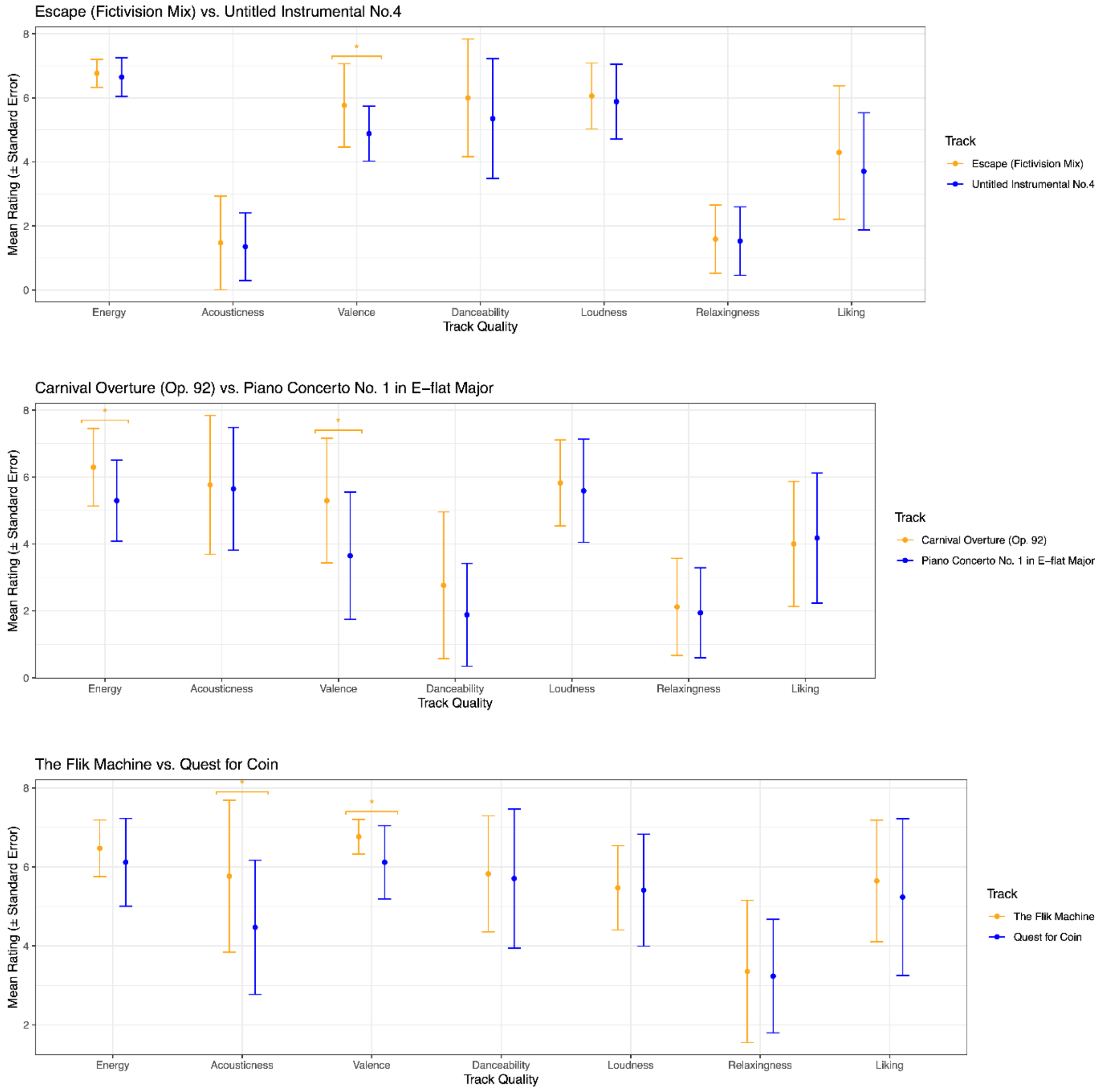


Figure C.2.

Mean values (± standard error of the mean) of track quality responses for low relaxation potential tracks compared within genres (top plot: electronic dance music genre tracks, middle plot: classical music genre tracks, bottom plot: jazz music genre tracks). Asterisks indicate differences that reached corrected statistical significance ($p < 0.007$).

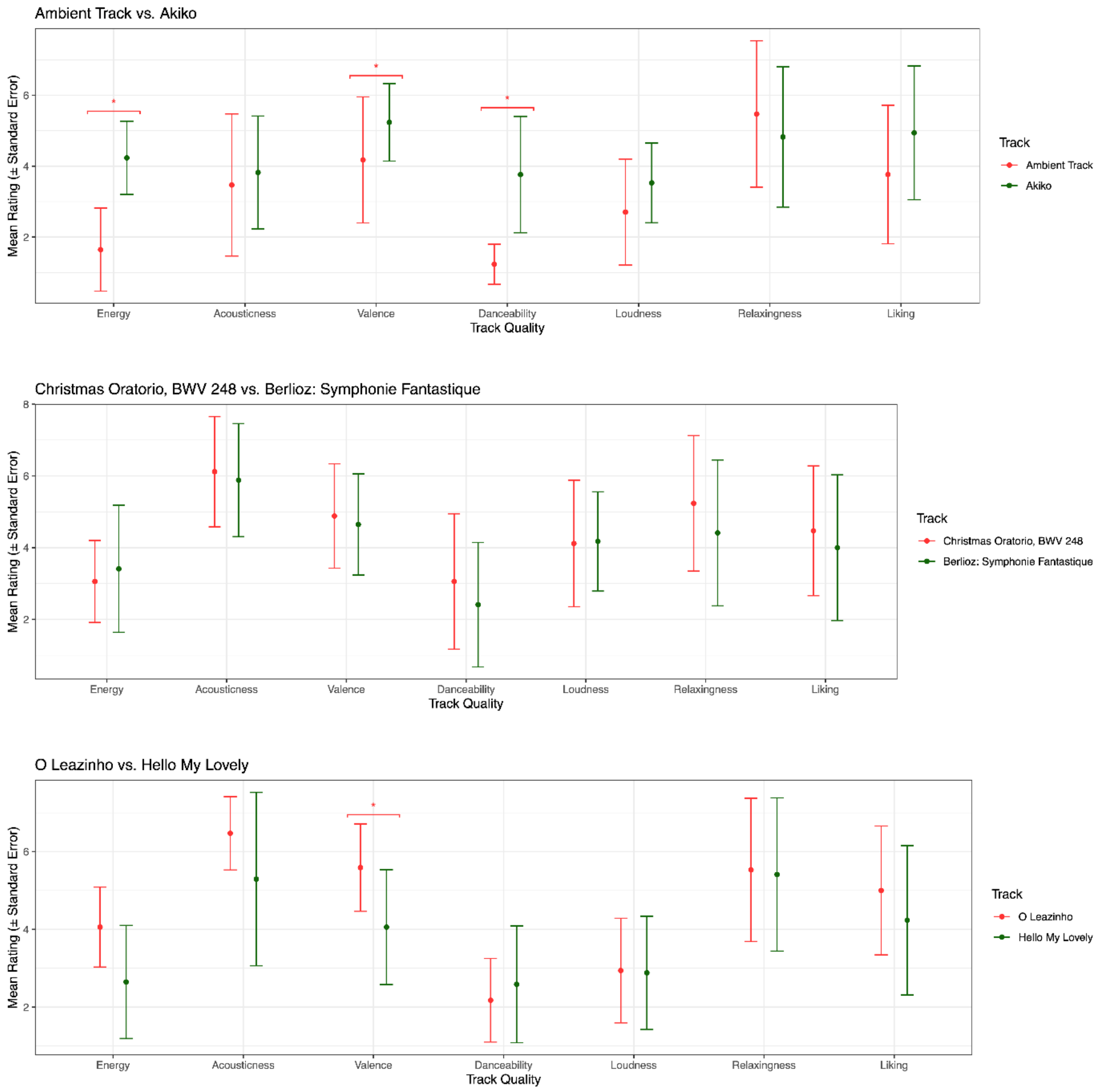


Figure C.3.

Scatterplot and regression lines (with banded standard error of the mean) of relationships between rate of visual imagery probes and each of the individual difference indices (musical training, Psi-Q, and STOMP) between the two categories of listening conditions: those with high or low Relaxation Potential (left) and high and low Familiarity (right).

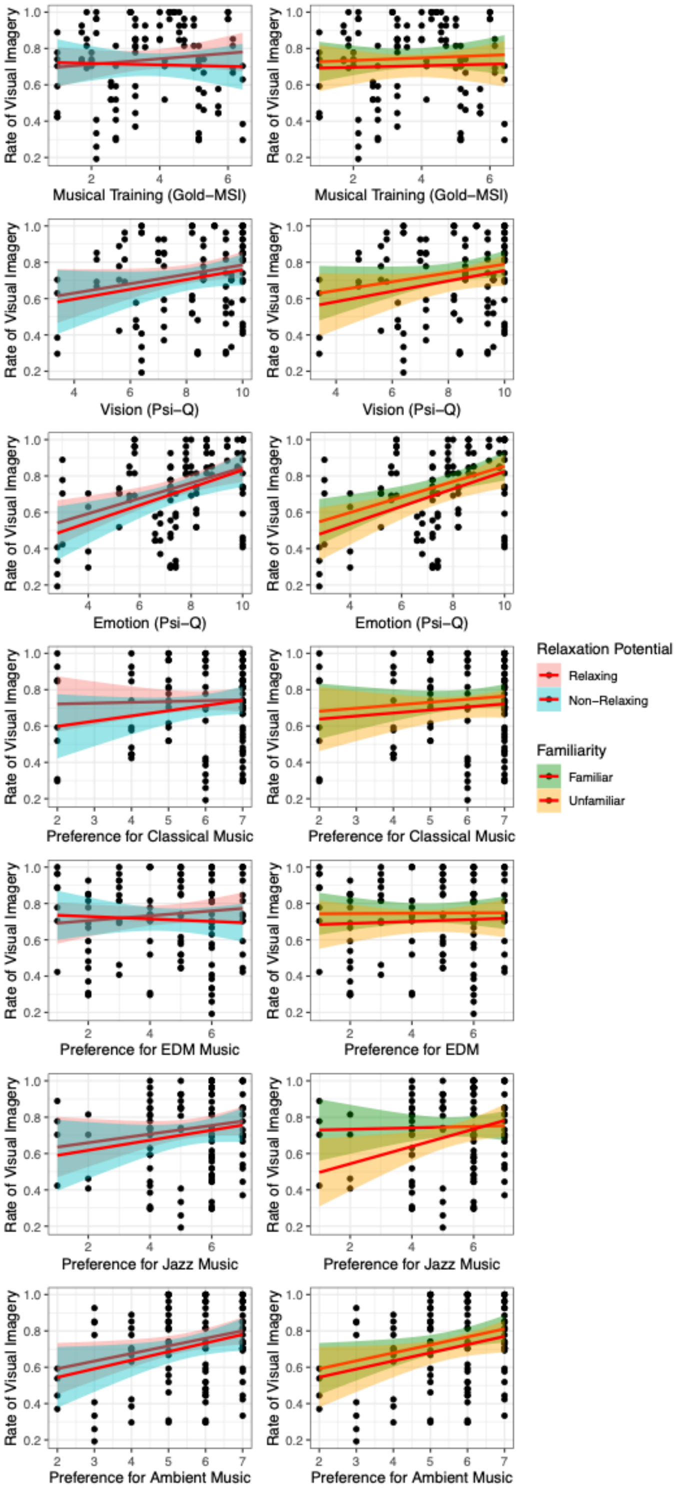
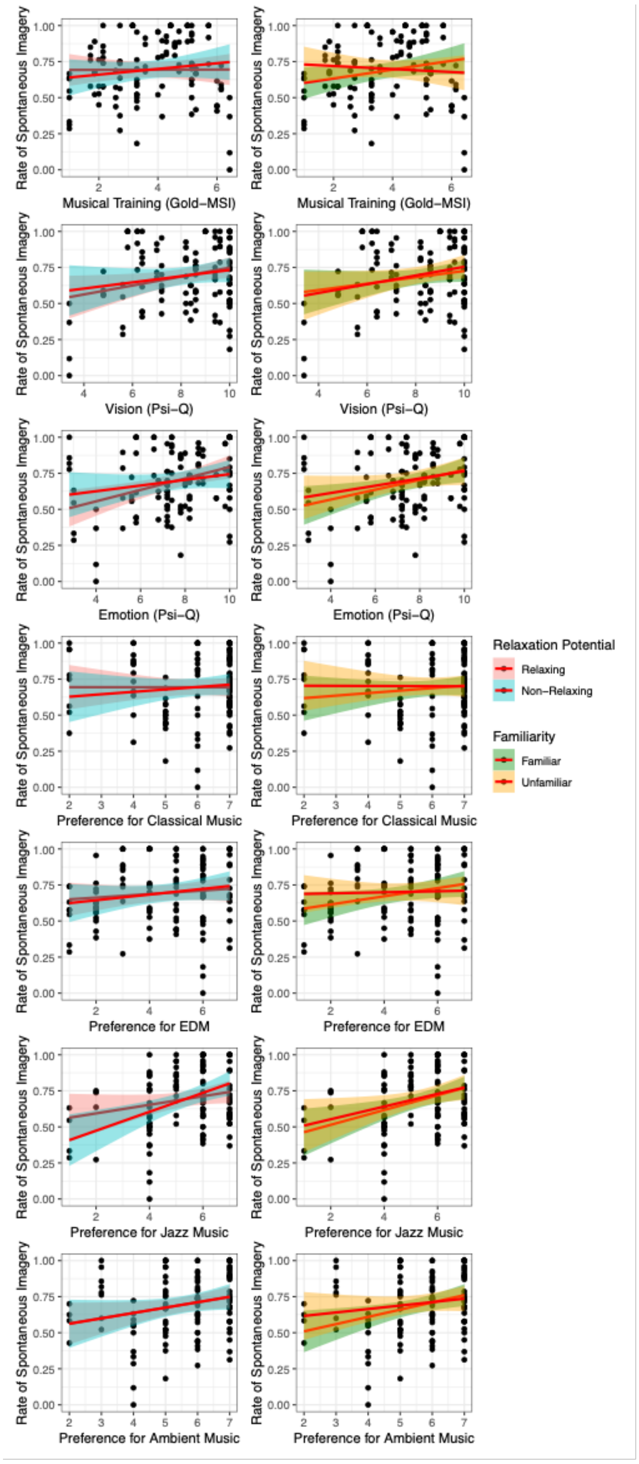


Figure C.4.

Scatterplot and regression lines (with banded standard error of the mean) of relationships between rate of spontaneous visual imagery probes and each of the individual difference indices (musical training, Psi-Q, and STOMP) between the two categories of listening conditions: those with high or low Relaxation Potential (left) and high and low Familiarity (right).



Appendix D. Chapter 6 Supplementary Tables and Figures

Table D.1.

Differences in behavioural measures between pure and minimal aphantasics and controls.

	***Pure* - Controls**						***Minimal* - Controls**					
	ß	**SE**	**95% CI**	***t***	***p***	**Partial η^2**	**ß**	**SE**	**95% CI**	***t***	***p***	**Partial η^2**
Prevalence	1.07	0.08	[0.60, 1.00]	13.49	**< .001***	.69	0.98	0.11	[0.43, 1.00]	9.22	**< .001***	0.56
Vividness	1.06	0.09	[0.56, 1.00]	12.11	**< .001***	.66	0.96	0.11	[0.39, 1.00]	8.60	**< .001***	0.53
Liking	0.04	0.14	[0.00, 1.00]	0.283	.778	.00	0.18	0.19	[0.00, 1.00]	0.95	.348	0.01
Felt Emotional Intensity	0.46	0.21	[0.00, 1.00]	2.13	.036	.05	0.80	0.25	[0.03, 1.00]	3.15	**.002***	0.13

Note. * significant at an alpha level of 0.013.

Figure D.1.

Density plots of the four behavioural measures (prevalence and vividness of visual imagery, music liking, and felt emotional intensity) between pure aphantasics and controls.

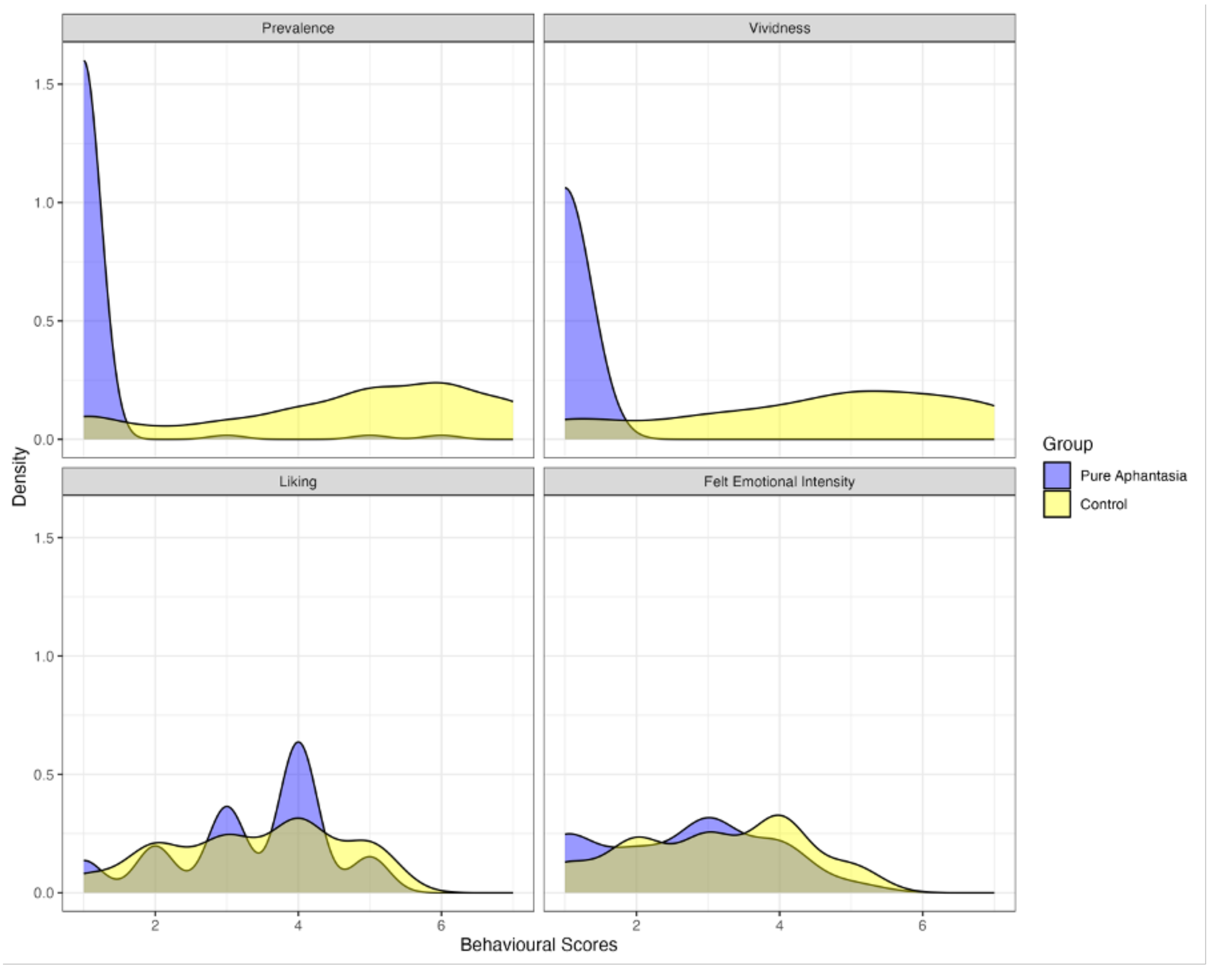


Figure D.2.

Density plots of the four behavioural measures (prevalence and vividness of visual imagery, music liking, and felt emotional intensity) between minimal aphantasics and controls.

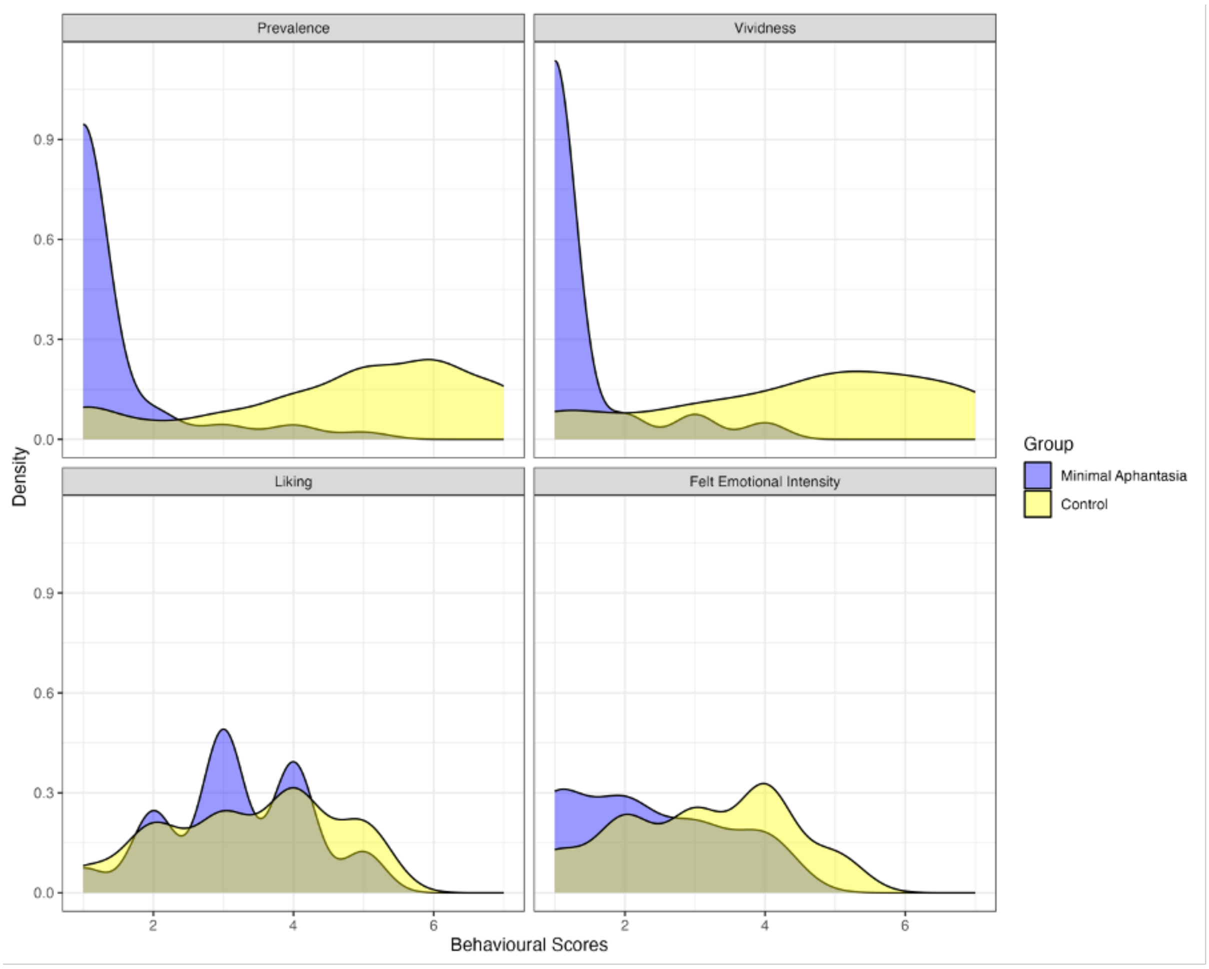


Table D.2.

Original MecScale items and adapted MecScale items.

Original	**Modified**
Did the music feature an event that startled you?	To what extent do you find that music features events that startle you?
Did the music have a strong and captivating pulse/rhythm?	To what extent do you find that music can have a strong and captivating pulse/rhythm?
Did the music evoke a memory of an event from your life?	To what extent do you find that music is able to evoke memories of events from your life?
Did the music evoke more general associations?	To what extent do you find that music is able to evoke general associations?
Did the music evoke images while you were listening?	To what extent do you find that music is able to evoke images while you are listening?
Were you touched by the emotional expression of the music?	To what extent do you find that music is able to touch you through its emotional expression?
Was it difficult to guess how the music (e.g., melody) would continue over time?	To what extent do you find that music keeps you guessing as to how it will continue over time?
Did the music have any practical consequences for your goals or plans in life?	To what extent do you find that music seems to have practical consequences for your goals or plans in life?

Note. Original scale items taken from the study by Juslin et al. (2014) were rated in a dichotomous fashion (Yes/No). The modified item used here were rated on a 7-point Likert scale, from 1 (Strongly Disagree) to 7 (Strongly Agree).

Table D.3.

Fixed effect estimates of multivariate linear mixed model on MecScale items between the ***pure*** *aphantasic and control groups.*

	ß	**SE**	**95% CI**	***t***	***p***	**Partial ŋ²**
Brain Stem Reflex	1.06	0.51	[0.00, 1.00]	2.09	.043	.10
Rhythmic Entrainment	0.30	0.39	[0.00, 1.00]	0.75	.456	.01
Episodic Memory	2.40	0.46	[0.21, 1.00]	5.19	< .001*	.40
Evaluative Conditioning	1.37	0.42	[0.05, 1.00]	3.26	.002*	.21
Visual Imagery	4.35	0.32	[0.74, 1.00]	13.72	< .001*	.82
Emotional Contagion	0.97	0.49	[0.00, 1.00]	1.99	.054	.09
Musical Expectancy	1.61	0.44	[0.08, 1.00]	3.63	< .001*	.24
Cognitive Appraisal	0.98	0.47	[0.00, 1.00]	2.06	.045	.09

Note. * significant at an alpha level of 0.006.

Table D.4.

Fixed effect estimates of multivariate linear mixed model on MecScale items between the ***minimal*** *aphantasic and control groups.*

	ß	**SE**	**95% CI**	***t***	***p***	**Partial η^2**
Brain Stem Reflex	0.18	0.68	[0.00, 1.00]	0.26	.799	.00
Rhythmic Entrainment	–0.83	0.40	[0.00, 1.00]	–2.05	.049	.12
Episodic Memory	0.39	0.43	[0.00, 1.00]	0.91	.370	.03
Evaluative Conditioning	0.29	0.43	[0.00, 1.00]	0.67	.506	.01
Visual Imagery	3.56	0.47	[0.46, 1.00]	7.50	< .001*	.64
Emotional Contagion	–0.18	0.45	[0.00, 1.00]	–0.39	.697	.00
Musical Expectancy	0.09	0.57	[0.00, 1.00]	0.16	.874	.00
Cognitive Appraisal	0.48	0.58	[0.00, 1.00]	0.83	.412	.02

Note. * significant at an alpha level of 0.006.

Table D.5.

Fixed effect estimates of multivariate linear mixed model on AFML items between the ***pure*** aphantasic and control groups.

	ß	**SE**	**95% CI**	***t***	***p***	**Partial ŋ²**
Stress Regulation	0.72	0.28	[0.01, 1.00]	2.53	.015	.12
Strong Emotional Experiences	1.07	0.30	[0.07, 1.00]	3.58	< .001*	.22
Rumination	0.46	0.30	[0.00, 1.00]	1.52	.135	.05
Sleep	0.11	0.37	[0.00, 1.00]	0.30	.766	.00
Reminiscence	1.32	0.26	[0.19, 1.00]	5.18	< .001*	.37
Anger Regulation	0.56	0.32	[0.00, 1.00]	1.76	.085	.06
Anxiety Regulation	0.48	0.31	[0.00, 1.00]	1.56	.126	.05
Awe & Appreciation	0.38	0.29	[0.00, 1.00]	1.30	.202	.04
Loneliness Regulation	0.49	0.32	[0.00, 1.00]	1.56	.126	.05
Cognitive Regulation	0.55	0.44	[0.00, 1.00]	1.27	.210	.03
Identity	0.29	0.31	[0.00, 1.00]	0.95	.348	.02

Note. * significant at an alpha level of 0.005.

Table D.6.

Fixed effect estimates of multivariate linear mixed model on AFML items between the ***minimal*** *aphantasic and control groups.*

	ß	**SE**	**95% CI**	***t***	***p***	**Partial ŋ²**
Stress Regulation	0.50	0.33	[0.00, 1.00]	1.52	.137	.06
Strong Emotional Experiences	–0.17	0.37	[0.00, 1.00]	–0.45	.654	.00
Rumination	0.21	0.33	[0.00, 1.00]	0.65	.521	.01
Sleep	0.09	0.49	[0.00, 1.00]	0.18	.859	.00
Reminiscence	0.47	0.31	[0.00, 1.00]	1.52	.137	.06
Anger Regulation	0.09	0.37	[0.00, 1.00]	0.24	.811	.00
Anxiety Regulation	0.15	0.37	[0.00, 1.00]	0.41	.687	.00
Awe & Appreciation	–0.16	0.34	[0.00, 1.00]	–0.46	.648	.00
Loneliness Regulation	0.19	0.37	[0.00, 1.00]	0.53	.601	.00
Cognitive Regulation	1.04	0.51	[0.00, 1.00]	2.07	.045	.10
Identity	–0.08	0.38	[0.00, 1.00]	–0.21	.831	.00

Note. * significant at an alpha level of 0.005.

Table D.7.

Fixed effect estimates of multivariate linear mixed model on BMRQ items between the aphantasic and control groups.

	ß	SE	95% CI	*t*	*p*	Partial ŋ²
Emotional Evocation	0.59	0.97	[0.00, 1.00]	0.60	.548	.00
Sensory-Motor	0.38	0.54	[0.00, 1.00]	0.71	.484	.00
Mood Regulation	1.52	0.96	[0.00, 1.00]	1.58	.119	.00
Musical Seeking	1.48	0.64	[0.01, 1.00]	2.33	.024	.09
Social Reward	–0.24	0.93	[0.00, 1.00]	–0.25	.797	.00

Table D.8.

Fixed effect estimates of multivariate linear mixed model on BMRQ items between the ***pure*** *aphantasic and control groups.*

	ß	SE	95% CI	*t*	*p*	Partial ŋ²
Emotional Evocation	1.42	1.08	[0.00, 1.00]	1.31	.196	.04
Sensory-Motor	0.78	0.61	[0.00, 1.00]	1.26	.213	.03
Mood Regulation	1.43	1.10	[0.00, 1.00]	1.30	.201	.04
Musical Seeking	1.01	0.68	[0.00, 1.00]	1.48	.145	.05
Social Reward	0.53	1.08	[0.00, 1.00]	0.49	.627	.00

Table D.9.

Fixed effect estimates of multivariate linear mixed model on BMRQ items between the ***minimal*** *aphantasic and control groups.*

	ß	SE	95% CI	*t*	*p*	Partial ŋ^2
Emotional Evocation	–1.00	1.27	[0.00, 1.00]	–0.78	.438	.02
Sensory-Motor	–0.37	0.73	[0.00, 1.00]	–0.51	.612	.00
Mood Regulation	1.69	1.22	[0.00, 1.00]	1.39	.173	.05
Musical Seeking	2.39	0.92	[0.02, 1.00]	2.58	.014	.15
Social Reward	–1.71	1.30	[0.00, 1.00]	–1.32	.196	.04

Table D.10.

Fixed effect estimates of multivariate linear mixed model on Gold-MSI items between the aphantasic and control groups.

	ß	SE	95% CI	*t*	*p*	Partial ŋ^2
Active Engagement	0.31	0.30	[0.00, 1.00]	1.04	.304	.02
Perceptual Abilities	0.06	0.34	[0.00, 1.00]	0.18	.858	.00
Musical Training	–0.31	0.43	[0.00, 1.00]	–0.72	.473	.00
Singing Abilities	–0.08	0.39	[0.00, 1.00]	–0.21	.831	.00
Emotion	0.68	0.32	[0.00, 1.00]	2.11	.039	.07

Table D.11.

Fixed effect estimates of multivariate linear mixed model on Gold-MSI items between the ***pure*** *aphantasic and control groups.*

	ß	SE	95% CI	*t*	*p*	Partial ŋ²
Active Engagement	0.62	0.34	[0.00, 1.00]	1.84	.073	.07
Perceptual Abilities	–0.12	0.38	[0.00, 1.00]	–0.32	.751	.00
Musical Training	–0.61	0.50	[0.00, 1.00]	–1.21	.234	.03
Singing Abilities	–0.10	0.43	[0.00, 1.00]	–0.23	.820	.00
Emotion	0.70	0.36	[0.00, 1.00]	1.91	.062	.07

Table D.12.

Fixed effect estimates of multivariate linear mixed model on Gold-MSI items between the ***minimal*** *aphantasic and control groups.*

	ß	SE	95% CI	*t*	*p*	Partial ŋ²
Active Engagement	–0.28	0.43	[0.00, 1.00]	–0.66	.516	.01
Perceptual Abilities	0.41	0.45	[0.00, 1.00]	0.90	.373	.02
Musical Training	0.26	0.59	[0.00, 1.00]	0.44	.666	.00
Singing Abilities	–0.06	0.56	[0.00, 1.00]	–0.10	.920	.00
Emotion	0.64	0.42	[0.00, 1.00]	1.55	.131	.06

Figure D.3.

Averaged scores of the MecScale items of the ***pure*** *aphantasic and control groups. Error bars reflect ± standard error of the mean.*

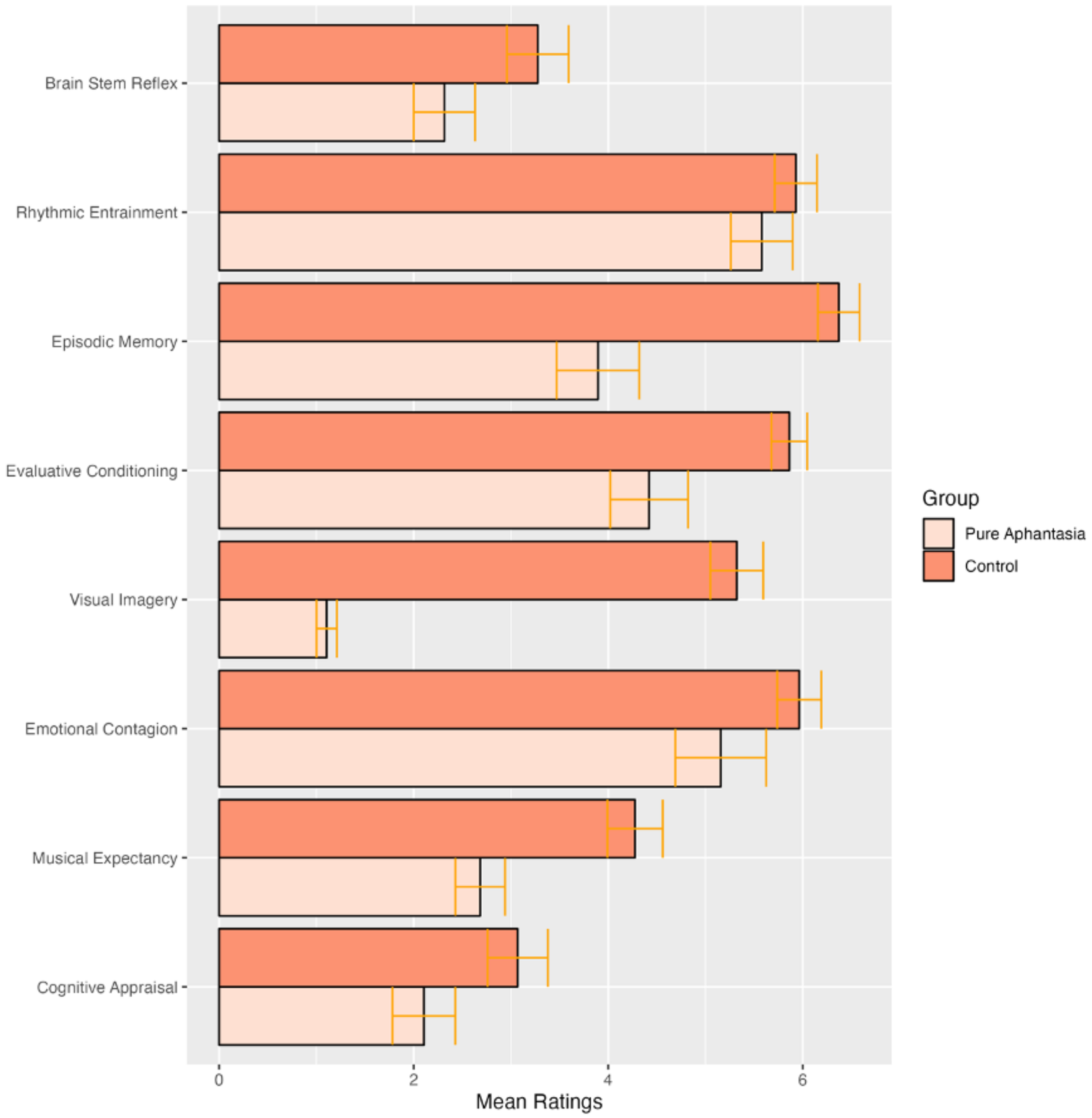


Figure D.4.

*Averaged scores of the MecScale items of the **minimal** aphantasic and control groups. Error bars reflect ± standard error of the mean.*

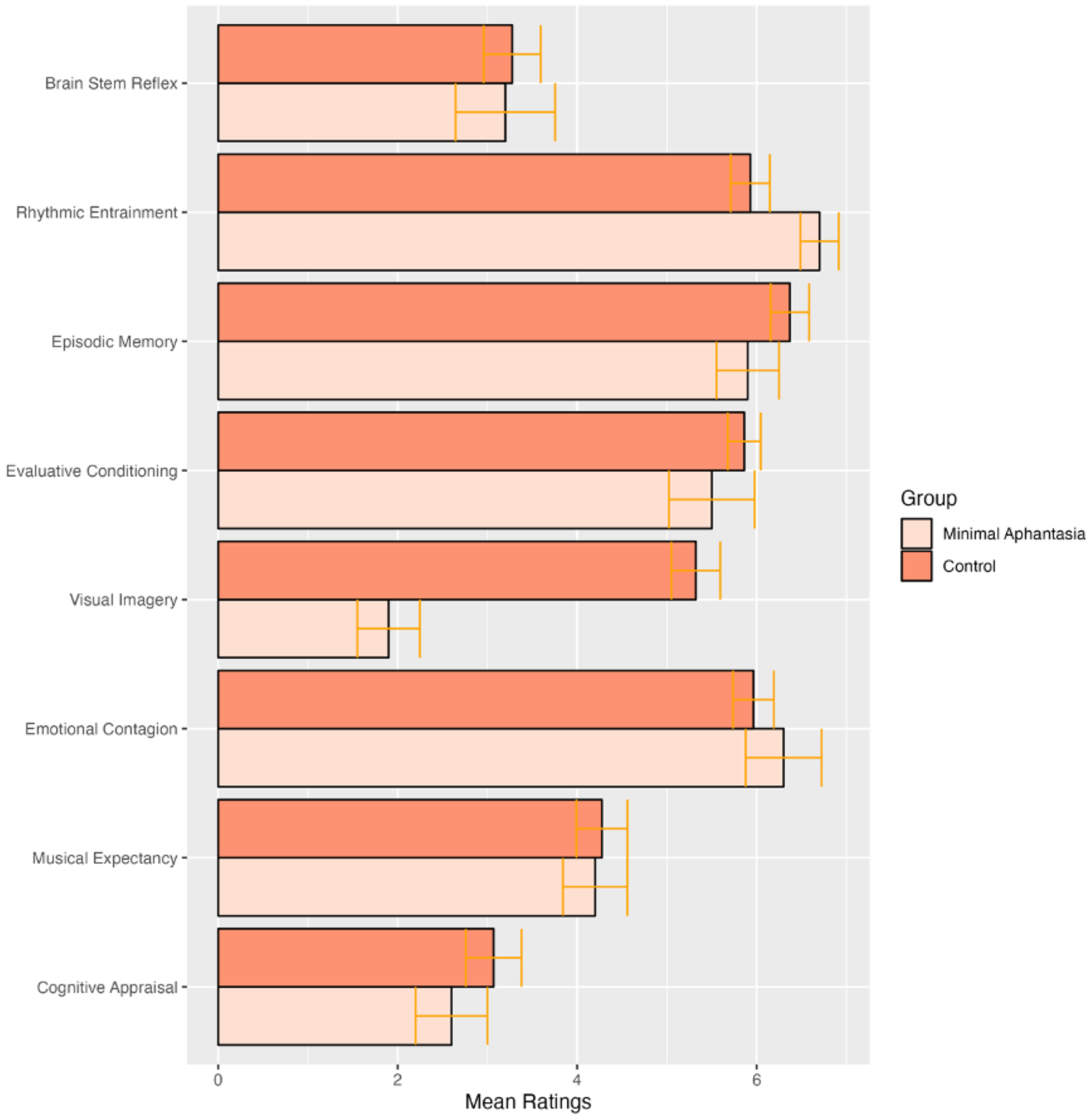


Figure D.5.

Averaged scores of factors for the Adaptive Functions of Music Listening Scale (AFML) of the ***pure*** *aphantasic and control groups. Error bars reflect ± standard error of the mean.*

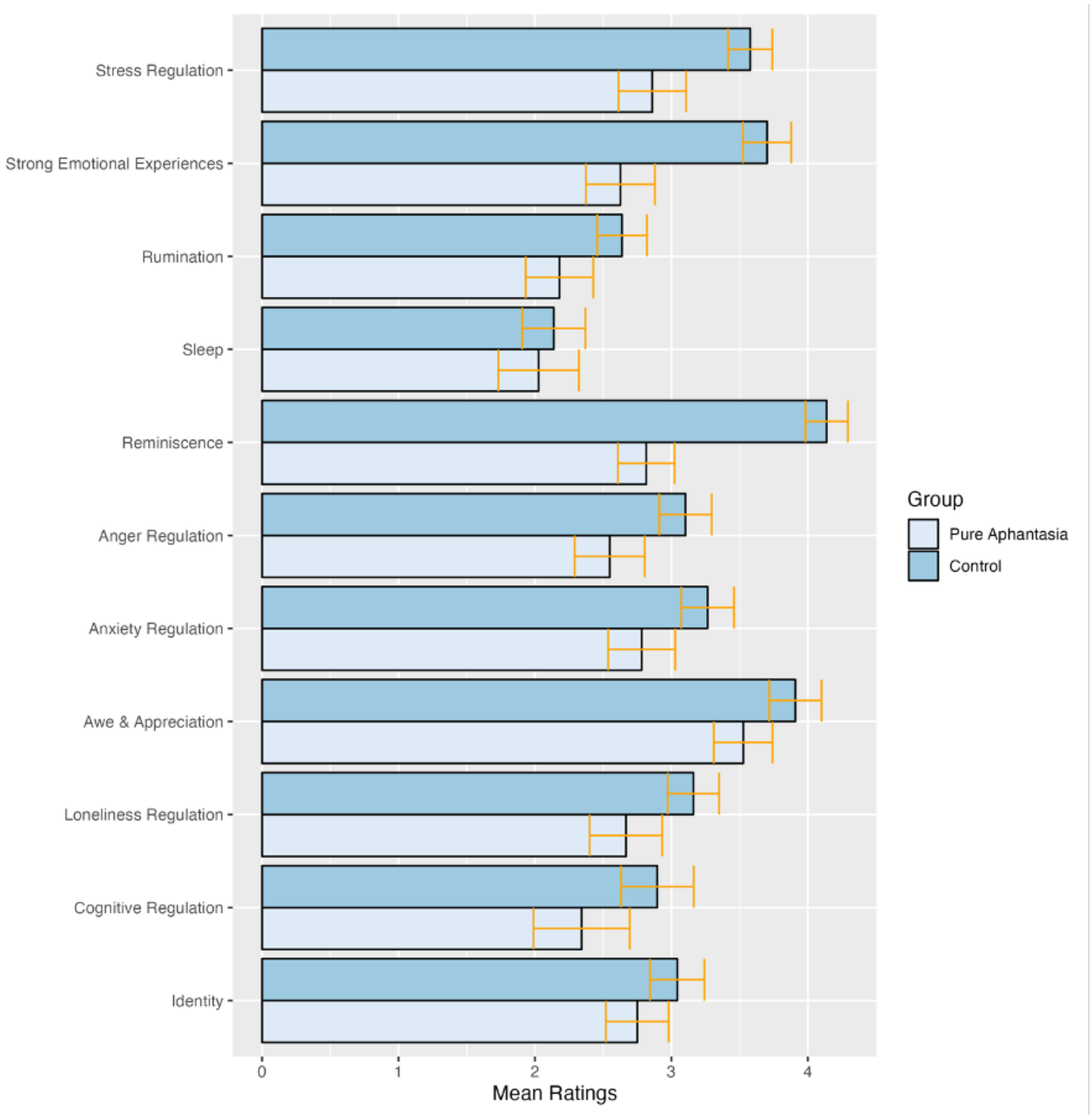


Figure D.6.

*Averaged scores of factors for the Adaptive Functions of Music Listening Scale (AFML) of the **minimal** aphantasic and control groups. Error bars reflect ± standard error of the mean.*

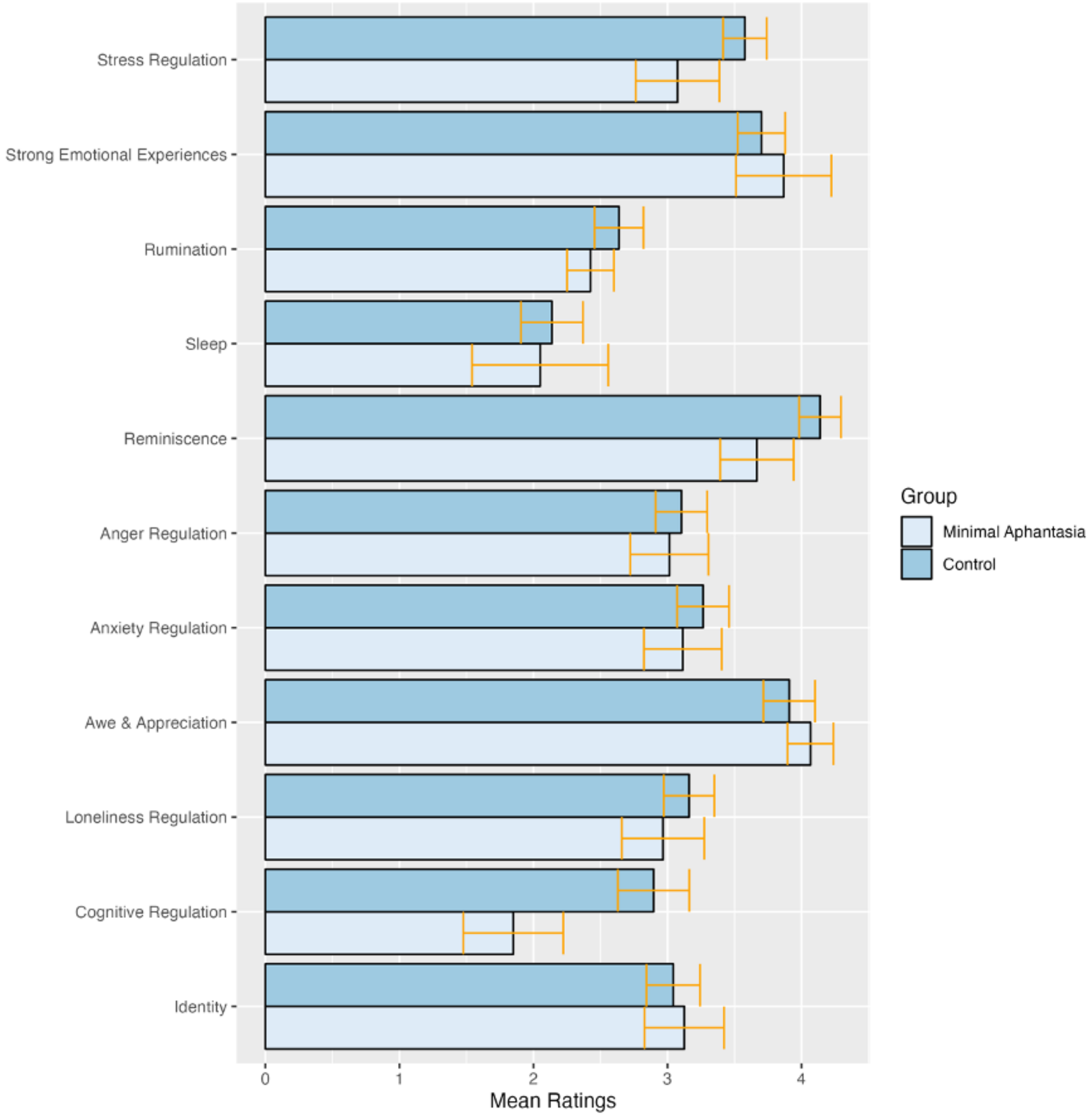


Figure D.7.

Averaged scores of factors for the Barcelona Music Reward Questionnaire (BMRQ) of the aphantasic and control groups. Error bars reflect ± standard error of the mean.

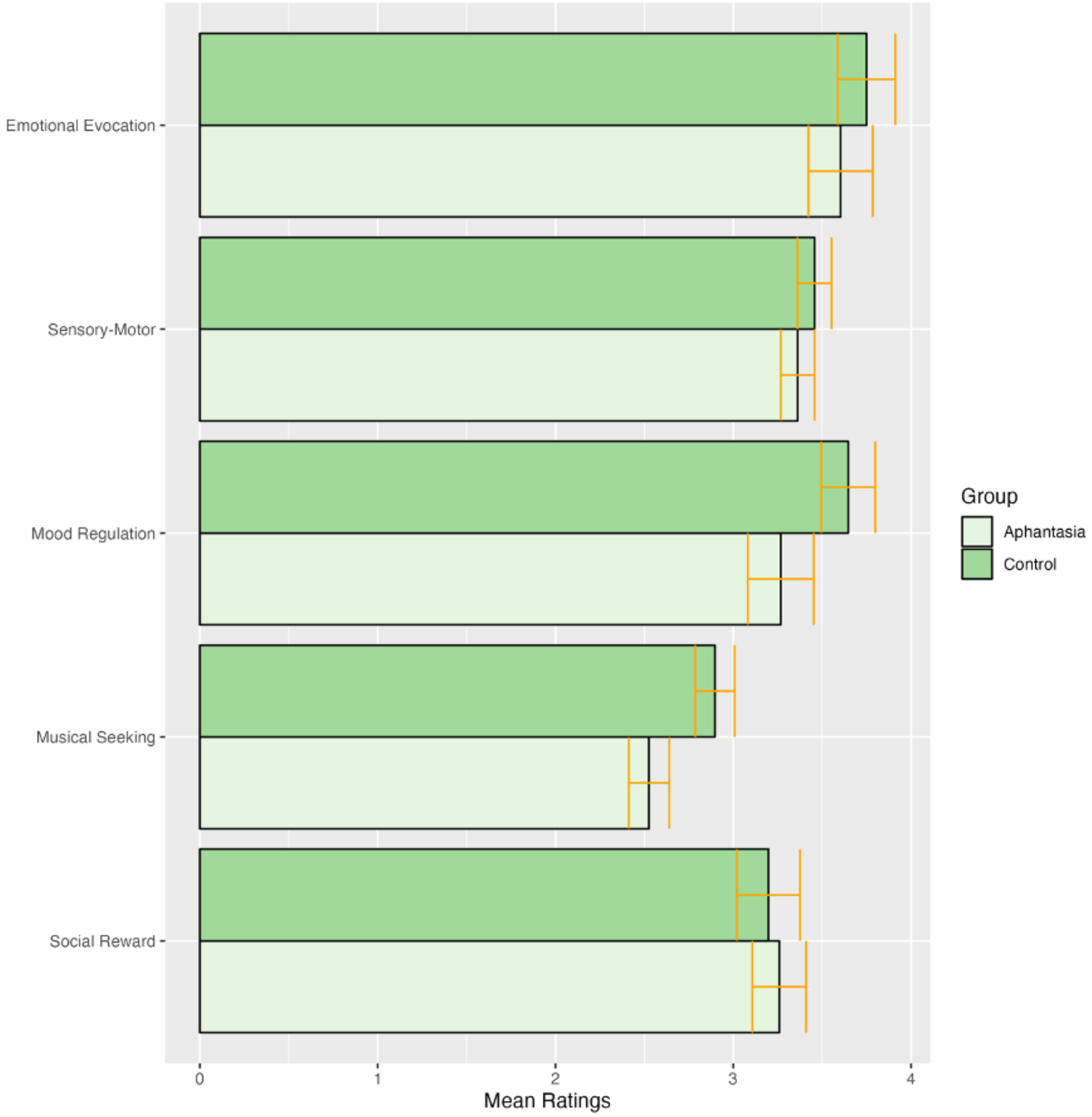


Figure D.8.

Averaged scores of factors for the Goldsmiths Musical Sophistication Index (Gold-MSI) of the aphantasic and control groups. Error bars reflect ± standard error of the mean.