University of Oslo:
Department of Music
ology

Norwegian University of
Science & Technology:
Department of Music

Cross-modal correspondences:

Different modes, common codes?

Investigating musical engagement with an Ecological cognitive approach

JONI MOK

Master's programme in
Music, Communication and
Technology

May 2022

Abstract

This project uses music to explore the way in which the environment affects cross-modal correspondences (CMC). This phenomenon relates to the way senses might be synchronised. Gibson's ecological theory of perception and the subsidiary theory of 4E cognition provide the theoretical starting points for the research. The thesis builds on data collected in relation to the public research concert MusicLab 8: Synaesthesia. The thesis consists of a two-stage experiment, one representing the live performance and another a home listening context. Data collection was done using open-ended surveys, interviews and audio-visual recordings. The data is studied using qualitative methods such as content analysis and coding. The significance of the work contributes to the multisensory experience in music and HCI research. The exploration of a 4E cognition framework opens up interesting possibilities to reconsider the generative role of musical engagement. With this in mind, it may provide practitioners with a more solid, holistic foundation when planning and designing musical events where cross-modalities are more important (i.e. audiences' participation, staging, sound array positioning, visualisation).

Acknowledgements

The work described in this thesis is based on participation in a collaborative research project. Every effort has been made to identify the specific contributions of colleagues in the text but in the nature of scientific research, ideas can be hard to always track and assign. I hope that nonetheless the source of contributions is adequately indicated.

A manuscript describing the research presented in this thesis has been submitted to the NordiCHI 2022 conference (late April 2022).

Thank you to **Alexander Refsum Jensenius** and **Laura Bishop** for being super-wise, helpful and patient supervisors who guided me all the way through, despite the fact that I might come up with broad ideas at times. Thanks to **Olivier Lartillot**, for sharing his computer science expertise and the opportunity to collaborate with his ingenious MIRAGE project. MusicLab, for the very interesting collaboration and learning opportunity to practice research design. Thank you to MCT fellow teachers and fellow students.

Penultimately, thanks to all the experiment participants and interviewees for their time providing input voluntarily.

Last but not least, I want to thank my husband, Richard and my inner circles: my parents and close friends for all their love and support down the years.

Word count: 17,313 (excl. Abstract, Acknowledgements, References and Appendices)

As required, a MCT blogpost link was included in Appendix 8.

Contents

CHAPTER 1: INTRODUCTION

PERSONAL MOTIVATION

Since I was a child, I was known for my unusually good memory, especially people's birth dates and numbers. Even today, I can still remember Henry's birthday (we met in 1st grade) - it's November 18. These kinds of meaningless numbers have remained in my head for two decades. Is having good memory for numbers actually good for me at all? I ask myself this once in a while and the answer is usually: "It depends". Whenever a number is written or spoken and if this information is known to me, immediately, mentally, I see where the number could fit into my "imagery grid system". This puzzling "mental imagery" kept me wondering for years. It also inspired me to easily create and connect completely different things. It also led me to have a strong interest in reading cognitive psychology literature as a teenager. With my rather nerdy interest and curiosity, I once thought that it was impossible to combine cognitive psychology with my occupations as a musician and designer. This lost dream emerged again five years ago when I was asked to design something that has meaning to me. I started to dive deep into the topics of synaesthesia, attention and memory and I designed prototypes (Mok, 2017). This discovery and curiosity led me further down the path of designing multisensory interactions for people who may need them. In 2018, I worked on research projects such as speech therapy and explored how images could trigger sounds, in order to assist young children (age 3 - 5) in pronouncing Danish better. I also explored how the use of colour and typefaces (shapes of letters) could improve dyslexics' reading experience (Mok, 2018). Because of these new discoveries and learning, I began to see a connection between cognitive psychology, design and music. In this thesis, I want to explore from a more theoretical perspective what triggers our thinking and feeling via the different stimulation of our senses. Can this current study lead to something even more meaningful? Can using multisensory interaction design improve the lives of others? Or simply to help people learn better? Can music, design and cognitive psychology be related to creativity (Herriott and Mok, 2021)? These are some of the ingredients making up my personal motivation for this project.

IMPACT OF THIS STUDY

As a student in the Master´s degree in Music, Communication and Technology, I have been particularly interested in the second part, the communication link between music and technology. Music is nothing if it does not communicate. Technology fails if it gets in the way of the message. In my research, I wanted to bring a 4E cognition approach to inquire into human perception from an ecological perspective to support communication. As we will get back to later, the four E's stand for embodied, embedded, enactive, and extended cognition. It is desirable to have surroundings which support the message of the music. Music has multi-modal effects and my work shows, albeit in a small way, that the way we listen to music affects its perceived content. Staging (or the context of the music) should at least support the music´s content; we could find out more about how to enhance its content; we could find out how to reduce distraction and ensure a greater chance the audience shares the same experience; we can also begin to analyse why well-staged performances of the past worked as they did.

Human-Computer-Interaction (HCI) involves a combination of channels: haptics, visual, and auditory. At present, the way in which the modes interfere or support one another is not considered. Understanding cross-modal interaction is crucial for better HCI development.

INTRODUCTION TO HCI, MUSIC LISTENING AND CROSS-MODAL CORRESPONDENCES

“The only information that reaches us concerning outward events appears to pass through the avenue of our senses, and the more perceptive the senses are of difference, the larger is the field upon which our judgement and intelligence can act”. - Francis Galton, noted in 1883 (Spence, 2021: 13)

To reduce human-computer-interaction´s (HCI) over-reliance on blinking, flashing lights and abrasive colour combinations, other sensory channels could be exploited e.g. sound. But this must be done with care. Humans live in an environment that they constantly need to make sense of. This means, among other demands, that inputs to their senses need to be coordinated. If the integration of senses were incongruent, it could be distracting and hazardous. Burke (1978, vii) describes the causes and interconnections of technological failure due to the interdependence of nature of humans and technology: “We live surrounded by objects and systems that we take for granted, but which profoundly affect the way we behave, think, work, play, and in general conduct our lives and those of our children”. Visual and auditory channels are targeted by technology because of the ease with which large amounts of information can be communicated to the user. Such targeting exploits the “higher rational senses, our hearing and our vision” (Spence, 2021, p.17). Our sensory channels are the avenues by which we are affected and it is necessary to reconsider how those channels interact with each other.

We communicate with others through the use of electronic devices, we listen to music while travelling or walking the dog. Here we might begin to ask some questions: does our listening experience get affected by the way we pay attention as much as the actual event itself? If there were a nagging couple arguing on the street in front of us while we are listening to upbeat, happy-sounding music, would the music affect our understanding of what we see in our immediate environment? How does the colour of the room affect the disembodied music we hear on our Marshalls? These questions of cross-modal binding relate to the communication of abstract information, such as emotion and the meaning of a given situation to perceivers. Through a better understanding of the biases of human perception, we could help reduce the problems technology may cause. We could reduce the use of blinking, flashing, and abrasive icons. HCI could be not only more effective but also simply nicer.

In HCI research, the matter of how to communicate information to users has been an important object of study. With the invention of the graphical user interface (GUI), information can be conveyed to users visually through the use of colours and icons. The use of auditory stimulus in HCI has historically been used to communicate messages to users through short non-speech audio sounds (Bramwell-Dicks et al., 2013). However, can a beeping sound or a flashing green light array become more personally meaningful? Instead of non-speech audio sound, recent research interests expand to emotion and music perception. Could more be done with sound than is currently done and what would the issues be in so doing?

This thesis builds on earlier research into music and HCI (Holland et al., 2013; 2016; 2019) and multisensory research in HCI (Obrist et al., 2017a; 2017b; Lin et al., 2021). It explores how visual and auditory information can mutually reinforce or modify one another, evoke emotions or elicit mental imagery. The present contribution offers an account of cross-modal correspondence in music listening that is grounded on exploration, and engagement with music through technology. The theoretical background was grounded on the ecological perceptual theory and its complementary framework known as the 4E cognition. I base such insights on the results of an original mixed-methods study with a guitarist and audiences in both live concert and home listening settings.

KEY TERMS IN THIS THESIS

Perceiver/listener/subject/participants: A person/subject who perceives/listens to/participates in something.

Stimulus (pl. stimuli): the source of sense information or sense-data which activates our senses (e.g. a guitar sound is a stimulus activating our auditory sense).

Congruent/ incongruent: compatible/incompatible; matched/mismatched e.g. if you see a moving vehicle up close but can hear no sound at all it is incongruent.

Touchpoint: A communication touchpoint is a meeting point where both the sender and receiver meet e.g it could be between humans and/or between humans and computers. There are two types of a touchpoint e.g. 1) a meeting room is a touchpoint for two people talking to each other and 2) when a person makes contact with a control on a machine such as a volume control slider.

Stakeholder: In a discussion about a proposed action, a stakeholder is anyone affected by or with an interest in the outcome.

Mental imagery: It refers to representations and the accompanying experience of sensory information without a direct external stimulus. It is understood to be multimodal.

Association/mapping/correspondence: The correlation between the effect of one stimulus and the effect of another stimulus e.g. loud sound is normally associated with bright colour.

Affordance: Affordance is a feature in the environment that a person knows what to do with or how to avoid.

Note: Rather than use the pronouns he/she, I have elected to use the third person plural "they" e.g. "when a person hears a sound they might enjoy the experience".

1.1 Aim, Objectives and Research questions

This study explored: ***(RQ0) How is music perception affected by the sensory channel a subject attends to?*** Since the nature of this question is broad, I will break it down into 3 sub-questions as below:

RQ 1: What is the relation of the listening conditions to the way the music content is understood in a concert setting?

RQ 2: What is the relation of the listening conditions to the way the music content is understood in a non-concert setting?

RQ 3: What is the difference between these two sets of results (RQ1) and (RQ2)?

1.2 Scope and Limitations

In this master's thesis, the aim is to investigate cross-modal correspondence from a communication[1] and user-experience[2] designer's perspective. Despite the fact that I am a classically trained flautist (I specialised in 20th-century French classics), this music performance background does not have much impact on the project. In terms of the approaches to this project, the project draws mainly from the social science methodologies such as the use of qualitative research.

For this thesis, the concepts of cross-modal correspondence (CMC) and situated listening experience were analysed in relation to theories of Gibson´s ecological theory of perception. In this case, I refer to qualitative analysis with relevant quantitative data upon which the findings are based. Naturalistic/statistical approaches, such as the use of mixed-method analysis (e.g. signal/acoustic analysis) will not be considered (though I will come back to them in the discussion section)[3].

1 Communication design aims to develop information and convey messages through multimedia. Its focus is on the communicative process and information development.

2 User experience design focuses on the design of socio-technical systems where people and computer systems are involved. UX design aims to organise the complexities of human nature and artificial information in order to make a better user experience happen.

3 I would like to add a note about how the time for this project was allocated. One day in November 2021 was spent collecting Phase 1 data, the concert. And 2 weeks to design the Phase 1 experiments and survey questions. The thesis construction was a set of overlapping activities which began in January 2022 and was concluded in May 2022. During the 4 months (equivalent 30 ECTs), I analysed data from phase 1 via NVivo using content analysis and coding methods. Meanwhile, I was planning and designing interactions, and recruiting participants for phases 2a and 2b data collection. I interviewed and transcribed the interviews then I collected more data that were comparable with phase 1. I coded and conducted content analysis another time with participants' involvement so as to reduce the influence of my own bias. For phase 2b, the last data was collected on April 5, so I then started to evaluate my data and analyse the content. During the practical part of the thesis, I also researched the theoretical material. The whole 30 ECTs were spent on what I mentioned above, therefore, for more analysis and future actions, I did not have time for a mixed-method analysis such as conducting signal analysis for the music stimulus and so on.

1.3 Overview and structure of thesis

This thesis consists of 5 chapters. Chapter 1 introduces some terms that will be used throughout the thesis. It also provides a general overview of what the thesis is about. Chapter 2 looks into the theoretical background and previous research work done on cross-modal correspondence. It also explores what music perception is and approaches to it i.e. the ecological theory of perception. Chapter 3 contains an account of the methodology which consists of a 2-phase experiment based on the theory and its design methods. Chapter 4 presents data analysis and the analysis methods I have used. Finally, Chapter 5 discusses the implications of the study.

CHAPTER 2: BACKGROUND

In this chapter, I will review four areas relevant to the subject of inquiry. First, I will look at (1) cross-modal correspondences (CMC) and the related concept of (2) synaesthesia. Then I will briefly outline the main points of (3) ecological perceptual theory; then I introduce (4) the complementary framework of cognition known as 4E. I introduce them in this order because the focus is on the concepts (1) and (2) in relation to the well-known general level theory (3) and the complementary framework (4).

2.1 Cross-modal correspondences

When we see an image of a hamburger in a fast-food restaurant, we are prompted to think of the taste of burgers we have previously eaten. The sound of a particular musical instrument might make us think of round shapes. In both these instances, sense experiences from one channel provoke sense impressions from another channel. More precisely, cross-modal correspondences (CMC) are the non-arbitrary perceptual associations of stimulus features both within and across senses (Spence and Parise, 2012; Spence, 2011; Schmitz et al., 2021; Spence, 2020a). It is safe to say that CMC is a widely experienced phenomenon. It is something "the majority of people share between seemingly-unrelated stimuli presented in different sensory modalities" (Spence, 2020a: 6)". CMCs are made pre-reflectively, so people often are not aware of them and may be unable to explain why two features seem to pair together. These associations have been broadly categorised under the following two headings:

1. ***Intramodal correspondence (within one sense):***
   Spence (2020a) suggests the term intramodal correspondence refers to correspondence experienced within one sensory modality e.g. two features of sounds such as pitch and rhythm.

2. ***Cross-modal correspondence (across senses):***
   On the other hand, cross-modal correspondence refers to how the association of one sensory channel influences another sensory channel. Associating colours with music is an example. Adeli et al.´s (2014) study found that their participants associated soft timbres with blue, green or light grey rounded shapes; harsh timbres were associat-

ed with red, yellow or dark grey sharp angular shapes. This cross-linkage of sensory experience raises the question of whether or not there is any fundamental tendency to match dimensions (e.g. pitch, brightness) and features (e.g. colour, shape) across distinct sensory modalities, or to map one dimension of experience onto another. Or is such cross-modality ever viewed as being a positive or negative sensory experience?

In summary, CMC is a complex phenomenon of the linkage of different senses both with modes and between modes and between present experience and remembered experiences. The phenomenon of CMC has been explored in both art and science, however, in quite different ways which we turn to in the next section.

CROSSMODAL CORRESPONDENCES IN ART AND SCIENCE

In artistic research, artists and musicians create cross-sensory interactive experiences (McDonnell, 2020a; McDonnell 2020b; Gsöllpointner, 2015: Rudenko et al., 2019) to evoke synaesthesia (I will come back to this term shortly). Artistic works suggest questions about the sense of awareness of the complexity of cross-modal sensory associations; why and how human perceptions respond to the multiple sensory stimuli cross-modally cannot simply be understood without a deeper inquiry into the individual's perceptive experience. Artistic research may provide clues about unsuccessful or challenging art: if a stimulus does not match our expectations, the mismatch of a cross-modally correspondence would be defined by the individual as incongruent (McCormick et al., 2021) or what the layman might call "not very good" or "weird".

Turning to lab work, researchers observe the different neural regions using tools, such as functional magnetic resonance imaging, known as fMRI (Revill et al., 2014; Olivetti Belardinelli et al., 2004) map out the spatial and temporal relations during decision-making related to multisensory experimental tasks. Other measurement methods, such as the Stroop test (Stroop, 1935) are used to find out how efficiently subjects match given stimuli, such as colour and sound (Palmer et al., 2016). These methods and tools Deroy and Spence (2016: 33) define as statistical matching, as they are "grounded in the statistical regularities present in the environment". These observable clues can discover consistency over a period of time to confirm certain cross-modal correspondences. Deroy and Spence (ibid.) present the argument that statistical matching is a form of absolute matching since certain dimensions pair with another dimension. For example, these correlations can be found between pitch and size or pitch and elevation (Parise et al., 2014).

CROSSMODAL CORRESPONDENCES IN NATURE AND IN CULTURE

Cross-modal correspondences can be found in nature and in culture. Spence (2011; 2020a) and Palmer et al. (2013; 2016) distinguished four hypotheses (with related methodologies) to explain the correlations across senses. They are 1) *structural* and 2) *statistical*, both of which are natural in origin. The others are 3) *semantically-mediated relations* and 4) *emotional mediation*, both of which appear to be more subjective and culture-dependent.

Structural associations are a factor when pairs of dimensions have analogous neural coding systems in the brain, such as the perception of brighter lights and louder tones. *Statistical correspondences* arise from pairs of dimensions that are correlated in the physical, external world, such as larger-sized objects corresponding to lower-pitch tones. Examples would be the size of the resonating body of stringed instruments such as a bass and a violin or the sounds of different-sized balls being dropped creating certain cross-modal perceptions (Grassi, 2005).

Those two hypotheses focus on objective measurements that seem to be more naturalistic. However, the two hypotheses lack much by way of richness in terms of investigating the subjective, interpretative aspect that is experienced by individuals.

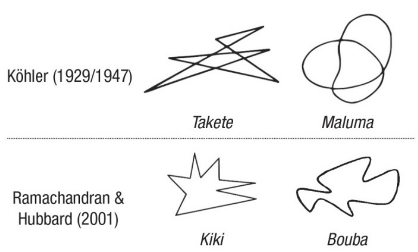


Figure 1. These examples (Köhler, 1929; 1947; Ramachandran et al., 2001) show the studies of semantically-mediated correspondences between images and word sounds.

There are other types of correspondences that are more abstract that cannot easily be explained by physical properties. For example, *semantically-mediated correspondences* arise from pairs of dimensions that have similar meanings. Ramachandran and Hubbard (2001) developed ideas which originated with Köhler´s (1929, 1947) interest in semantically-mediated correspondence. The 2001 work found a strong correlation between some images and word sounds. The nonsense word *kiki* suggested spiky, pointy things while the nonsense word *bouba* suggested a rounder object (see Figure 1 above). What might make us share the same cross-modal association when the stimuli are merely nonsense words and geometric shapes? Lakoff et al. (1980) proposed that these ways of crossing sensory domains could be understood as metaphors, that is to say, the sound was analogous to a metaphor for the geometry. Other examples of common metaphoric linkage included the fact that in almost all cultures "up" is positive and "down" is negative (Lakoff et al., 2003: 16). This metaphor argument rests on the idea that such mappings are in some way structurally inherent and intuitive. Zbikowski (2012) discusses how mapping between these domains would yield numerous possibilities for the sort of *meaning construction* associated with metaphor. Deroy and Spence (2016: 35) write that the intuitive character of these correspondences "seems to afford no further explanation – at the explicit level – than the fact that they just 'feel right'(…) Making sense of this intuitive aspect might be one of the challenges that lie ahead for researchers interested in the difference between correspondences and sensory metaphors". If feeling right is what makes most cross-sensory associations happen to individuals, what are the current inquiries that researchers are conducting into the subjective experience of individuals?

Spence (2020a; 2020b) and other researchers such as Palmer et al., (2016; 2013) write about the cross-modal relations in *emotional mediation*. This mediation involves subjective perception within individuals and concerns the dimensions of valences and activity. For example, Lin et al. (2021) provide evidence for the cross-modal correspondences between 3D tangible objects, colour brightness and emotional dimensions of the valence (pleasant/unpleasant) and arousal (excited/calm). Kawakami et al. (2013) discuss the correlation between minor-key music and the emotional dimension of valences. They found that their participants rated lower valance on minor-key music. There are also studies on how one relates stimuli to existing mental inventory, which would include memories (Watson et al., 2014).

ENVIRONMENT-AFFECT CORRESPONDENCES

Work has been done on what in plain terms are the relation of the environment to one´s senses and moods. Spence (2020b) points out that understanding our emotional responses to external stimuli can enhance our overall well-being. Brattico et al. (2019) find that aesthetic enjoyment can give individual music listeners an affective response through a musical engagement experience. Although much cross-modal correspondence research seems to focus on the quantitative aspects (Palmer et al., 2016; Fitzhugh et al., 2019; Olivetti Belardinelli et al., 2004; Langlois et al., 2013), in recent years, the use of qualitative analysis has become more popular (Lin et al., 2021). These later studies attempt to understand the subjective, emotional aspects of the correspondence.

To summarise this section, CMC can be experienced within one sense or across senses. These associations can be found in both nature and culture and there are related methodologies to investigate these mappings. Recent research investigates how individuals' emotional responses to external stimuli (multisensory environment) can affect one's overall well-being.

2.1.1 Synaesthesia

Many or indeed most researchers (e.g. Curwen, 2020; Harrison et al., 1996; Lacey et al., 2016) often mention both cross-modal sensory experience and synaesthetic experience simultaneously. However, in this thesis, I will treat cross-modal correspondence as being a broader, over-arching concept (see: Spence, 2011; Spence and Parise, 2016) or as Spence (2020a) called it, a *relative* correspondence that is experienced by general populations. Synaesthesia, in contrast, can be viewed as an *absolute* correspondence that refers to the specific neurological phenomenon experienced only by a distinct sub-group of the population (Cytowic and Eagleman, 2011; Ramachandran and Hubbard, 2001).

Synaesthesia is thought to be a form of (a sub-set of) cross-modal correspondence, understood in neurological terms. According to Baron-Cohen (1996) and Cytowic (1989), it is the experience of involuntary sensory cross-activation such that one stimulus influences another sensory stimulus to achieve a "union of the senses" (Cytowic, 1989). The most common types of synaesthesia are colour-music association (chromesthesia) and colour-letter association (grapheme-colour synaesthesia).

This phenomenon has been experienced by artists such as Wassily Kandinsky (Dreksler et al., 2019) and musicians such as Joni Mitchell (Espeland, 2006) who have made art using their synaesthetic experience. Other synaesthetes explored their own synaesthetic experiences by creating multisensory media (McDonnell, 2020a; 2020b; Rudenko et al., 2019). However, how are these sensory stimulations perceived by the audiences? Do they experience synaesthesia in the same way? How do the different sensory channels affect the way audiences perceive and experience?

Synaesthesia researchers also investigate different perspectives such as how auditory and visual information can mutually reinforce (Watson et al, 2014; Lacey et al, 2016), modify one another (Palmer et al., 2016; Langlois et al., 2013) or trigger one's mental imagery (Nanay, 2021; Nair et al., 2019).

To summarise this section, synaesthesia is a neurological phenomenon that researchers define as an *absolute* correspondence, experienced by a few, while cross-modal correspondence is defined as *relative* and is experienced by general populations. Many artists and researchers explore this concept by creating multisensory media and experiments.

In the next section, I will discuss the topic of music perception.

2.2 Music perception

Before we begin the discussion of music perception, it is necessary to ask what music is. Herbert (2013) offers the idea that music is information that does not require effortful decoding; its semantic ambiguity allows it to be easily customised by individuals. Noise (guitar feedback) and silence (John Cage) can be other forms of music depending on the context. They are the signals surrounding our everyday life. In this text, music is an organised sound we are in a position to consider, sounds created for aesthetic reasons. Of course, the topic is open always for further discussion.

Clarke (2005) sets up the reasons for using Gibson's ecological theory to explain musical meaning, by introducing philosophical approaches and 'information processing' approaches to music perception. Clarke emphasises that the act of listening is key to his theory of musical meaning. But at the same time, Gibson's theory does not rely only on the bottom-up processing hierarchy emphasised by most music cognition theories such as Temperley (2001) and Narmour (2000). Clarke explains that "music perception can be understood as a relationship between environmentally available information and the capacities, sensitivities, and interests of a perceiver".

MUSIC AS INFORMATION

Music, as information, can transmit complicated messages. It has content. So, when a tone is created, it creates vibrations in the air. Those air vibrations travel to a perceiver and affect the eardrums (Cook, 2001; Darrow, 1990), then the receiver´s brain processes those vibrations from the auditory nerves and transforms them into thoughts. The perceiver's understanding depends on the individual's ability to transmit that information or musical ideas across vast reaches of space and time. Humans are able to transmit knowledge across minds. A tone can be interpreted differently. Raindrops on the window create rhythm. Would some people perceive it as music? Does the way we interpret information shape our understanding of music?

MUSIC AS COMMUNICATION

In conversation, when someone´s feelings and ideas resonate with what another person just conveyed, we often hear the following phrases as a response:

> "We are on the same wavelength" or
> "I am in accord with you on that" or
> "your idea resonated with me".

The words *resonate with onesel*f, *in accord with* and *on the same wavelength* can initially be treated as a kind of phatic speech. However, it also raises an interesting question: why do they communicate this intersection as if they could see the "wavelength" or feel the signal is resonating with them? If music perception was a form of information transmission or a communication medium (Cross et al., 2009; Ruud, 1998), we could look at this as a whole communication system from the sender to the receiver's point of view (Umeozor, 2020; Shannon, 1948). However, what connects the two conversational partners such that they find themselves on the same wavelength and that they feel *resonated with one another?* Hargreaves (2012) explains this by referring to the idea of personal *musical geographies*; how we hear is filtered through and linked to our prior experiences, especially early ones. Listening is when we actively pay attention to what we are hearing. This active listening means linking what we hear to what we have heard. That store of ideas is our personal musical geography which is the mental map of one´s musical knowledge and experience in addition to one´s "cultural maps of musical reference" (ibid:.1). It is an embodied musical experience that mediates between music perception and action.

MUSIC AS A TRIGGER FOR CONSCIOUSNESS

Music perception can be described as a subset of the process in which an organism gains information from itself or the environment. The sense apparatus that have evolved for the natural world is directed toward music (active listening) or hearing it as sound (passively). Music perception has a cultural, learned element. Music perception can also refer to the state of awareness or the *act of becoming aware*. For example, a music performer, Petra plays a tune and the co-performer, Pablo listens to the tune and then improvises another tune as a response to the original. Petra may not know that she was playing in C minor because she did not pay close attention to her first improvised tune and believed it was C major, instead. She listens to the feedback from Pablo and becomes aware that it was in C minor. Her perception of the music is then altered based on the feedback she hears. Herbert (2013 and others) explore the idea that some aspects of perception can be unconscious. One could be distracted by other senses and often it might be that we may become aware of something without any knowledge of how. This could be understood by the use of phenomenological approaches (e.g. Høffding and Schiavio, 2021; Salice et al., 2019). Clarke (2005), on the other hand, presents a diagram to illustrate an idea of how sensory information is perceived from both a top-down and bottom-up manner:

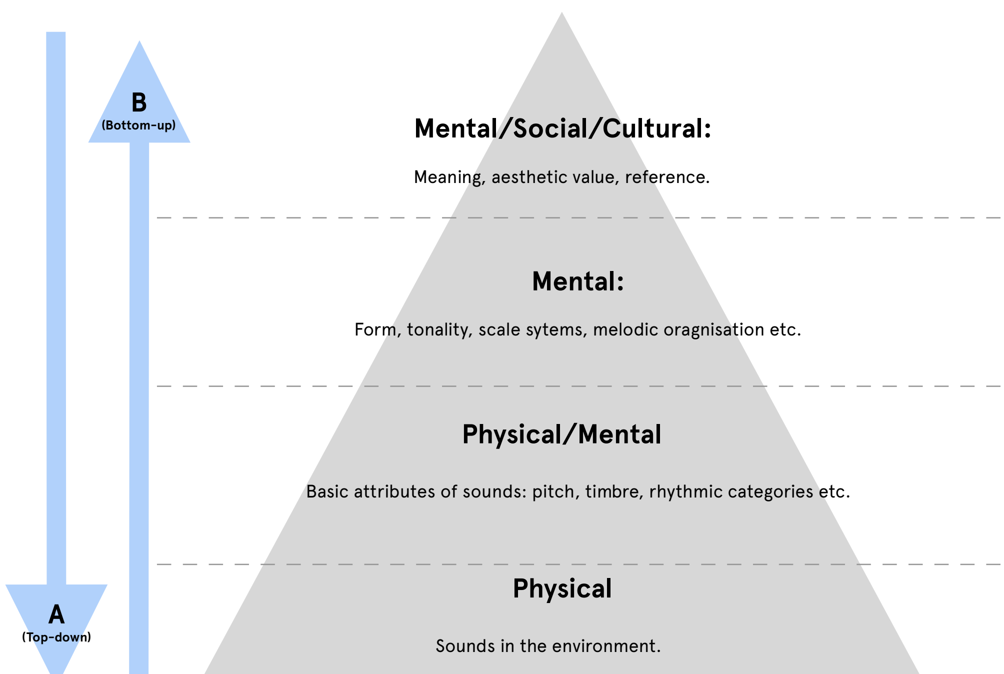


Figure 2. This diagram is adapted from Clarke (2005)'s Ways of listening: An ecological approach to the perception of musical meaning. I added two arrows that signify bottom-up and top-down processing in the diagram for clarity.

Figure 2 is adapted from Clarke (2005). It represents the four elements of the process of perception: 1) physical, 2) physical/mental, 3) mental and 4) mental/social/cultural, representing the possibilities of how human perceivers decode and encode information. *Arrow A* symbolises the "top-down" approach to music perception. It indicates how individuals perceive the environmental, and external signals. They process this information by using mental structures (mental model). Instead of analysing the signals of its objective properties, they use their existing mental frameworks to interpret the stimulus and construct judgement of *meaning, value and reference*. *Arrow B* illustrates the "bottom-up" approach to music perception. The environmental sounds are perceived by a perceiver. But this time, the perceiver processes the information from the physical level, unmediated. They make subsequent judgements based on the objective features of music, such as *pitch, rhythm and timbre*. Then, they begin to give meaning to these external stimuli. This bottom-up approach can be seen as computational. In parallel, highly trained music listeners with formal musical training may approach listening from a more analytical viewpoint such that personal emotion can be absent at first but later emerges. They could begin with the musical forms and the tonality before they start thinking about the emotional content of the sounds. However, contemporary lay listeners, almost all without formal musical training, tend to recognise the genres and the emotional content, such as the expressiveness of the performers before they start recognising the lower level musical properties such as the pitch of a music piece.

From the *physiological perspective*, the perceiver hears or sees the sensory signals such as pitch and brightness which trigger nerve pathways in the nervous system and lead to their perception in the conscious mind of the individual. A physical event is translated into nerve impulses and then into the subjective awareness of the person. Hearing involves pressure waves (Darrow, 1990; Cook, 2001). Vision involves light striking the retina of the eye (Wade et al., 2013; Shams et al., 2010). It can be seen as the organisation and interpretation of sensory information that later in the information-processing pathway is summed up as recognition of a change in the environment. But this kind of music perception seems to be a passive phenomenon: something happens outside the head that leads to something happening inside the head.

However, on the more *perceptual, cognitive level*, Clarke (2018) and many more researchers (Curwen, 2020; Bishop et al., 2018) argue that "there really is no such thing as passive listening". Music is a sound we listen to rather than just hear. Music can be understood as noise that we make a choice about. If we are not actively listening, then it is sound; music is necessarily sound we have attended to. Music is, as it were, the consequence of a choice the listener makes.

In this section, I have introduced the idea of how sensory signals can be picked up by individuals (see Figure 3 below). Depending on how the music fits with the individual perceiver's mental models, the sound can be experienced from a 1) more meaningful, value-centred "top-down" approach and a 2) more analytical, computational "bottom-up" approach. Music perception is therefore personal to individuals based on their musical geographies. Music can be treated as information and it can be communicated in a social setting. Music can also be seen as a trigger to our own consciousness that heightens our awareness.

In the next sub-section, I will briefly explore the correlation between music listening and colours. Colour is an element of vision and since this text deals with cross-modal correspondence, it is necessary for how auditory stimuli and visual stimuli relate to one another.

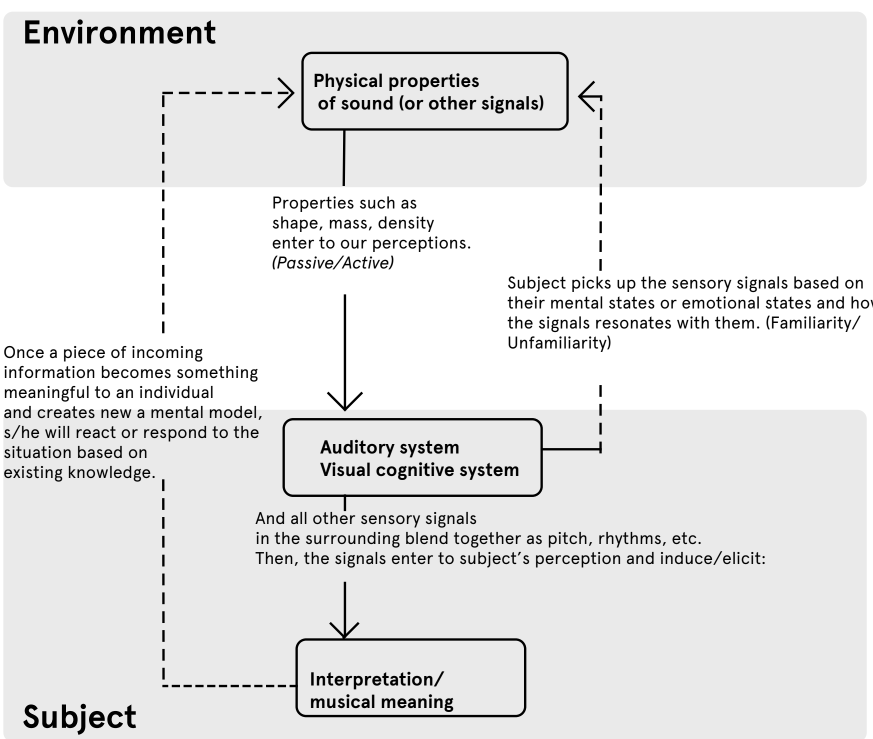


Figure 3. This diagram is intended to elaborate on my current understanding of how a subject picks up the environmental signals and processes the information. The environmental signals are picked up based on the individual perceiver's existing mental model. Once the new sensory signals are interpreted and become something meaningful to the individual, their mental model will be updated and they respond to the situation.

2.2.1 Colour-music correspondences

In this sub-section, I present an outline of findings dealing with the colour aspect of the visual sense in cross-modal correspondence, bearing in mind the two-way relationship between music and vision. For example, Huron (2015) describes how other modalities can affect the way music is perceived e.g. stage lighting can influence the audience's perception and emotional state. McDonald et al. (2022 :2) summarised some of the main findings regarding colour and we draw on them here in the following text. Colour may be well or less well matched to music with red being associated with high arousal (e.g. anger) and blue with low arousal (e.g. love) (Bresin, 2013). Palmer et al. (2013) looked into how music, emotion and colour were cross-modally associated. Yellow signified happiness while the low-arousal state of sadness was linked to blue. In Schubert et al.´s (2019) study slower music and or music in a minor mode was associated with bluer colours and sadder emotions.

These studies referred to in McDonald et al. (2022) indicate some linkage between colour (visual), music (auditory) and emotional states but presumably do not rule out creative matching and mismatching of colour and music, particularly for ambiguous and complex pairings. Movement and sound are cross-modally linked such that congruent sounds (e.g. an object moving left to right) will be detected faster if the sound also moves left to right (Meyer & Wuerger, 2001). Rhythm can also affect motion detection (Ten Oever et al, 2015). So, while we address colour, music and emotion, other factors can be involved that may not be so easy to separate out.

In the next section, I will introduce the ecological approach and its complementary framework of cognition known as 4E. This is intended to guide us to a more *holistic* picture of how musical engagement and other factors such as colour and emotions can be mapped.

2.3 Ecological approaches

AN ECOLOGICAL APPROACH TO MUSIC PERCEPTION

Gibson's ecological theory assumes that structure is inherent in the environment, not a construction within the mind (Hatfield, 2019). This conception focuses on aspects of the organism (animals, humans) and their surroundings that determine their behaviours. It is a contention that perceptual development is a gradual process of differentiating increasingly sophisticated levels of sensory stimulations. As Figure 2 in section 2.2 suggests, there are top-down and bottom-up approaches to how an organism processes the sense information from its surroundings. In the case of musical activity, ecological approaches can be used to analyse the 1) *social-cultural context*, which is more of the top-down processing and the 2) *environmental-natural context*, which is the bottom-up processing. Although both top-down and bottom-up processing most likely run simultaneously, the following paragraphs mean to examine the two processes one at a time.

In a *social-cultural context* of a musical event such as in an Irish pub, visitors understand that they can dance or act-as-they-wish during the live folk music. The pub performance is socially constructed. Visitors are encultured and share a mutual social understanding of the so-called "appropriate/standard" behaviours. In this musical context, the social norms of the "proper engagement" of how to listen and act may be determined by how the listeners interpret the sound, that is, "this sound means I can act this way". Put another way, we learn how to behave during listening and this, in turn, affects the way we process the sounds.

In an *environmental-natural context*, as the 20th-century composer Edgard Varèse, said "music is an organised sound". This saying indicated that when any sounds or noise in the environment were organised into some kind of structures or shapes, it could then be treated as music (this is dependent on the individual perceiver's preference). The modernist thinking of approaching music was to challenge the traditional understanding of music. For instance, the Western 12-tone scale music seemed to provide rules and had become a standard of what music could be. However, composers such as Olivier Messiaen, who was known for his birdsong-inspired composition such as *Le Merle Noir pour flûte et piano (The Blackbird)* wished to explore new structures. He turned environmental sounds, such as the birds' songs into organised sounds and so it became music. Varèse 's Density 21.5 explored the sound masses that were produced by a platinum flute with a density of 21.5 grams per cubic centimetre. The two compositions suggested that music can be recognised when noise or sounds in the environment were structured. A musical sound is then, determined by the physical properties of the producing instrument, such as shape, density and mass. This accords with an ecological theory which explains that it is pitch, rhythm, and instrument identification, not complex decoding of the stimulus that matters.

There are three factors that Clarke (2005) mentions that make the claim of ecological theory "both more realistic and more interesting":

1) perceptual learning,
2) adaptation and
3) the relation between action and perception.

Each of these 3 is discussed in the next sections.

Perceptual learning is what E.J. Gibson (1989) referred to when discussing processes that occur when organisms interact with the environment through sensory experience. It includes developing ability by perceiving, as well as performing a specific task. Gibson argued that knowledge is obtained through perception and that cognition is built on a foundation of perceptual knowledge. Despite the fact that there are opponents (Ullman, 1980) Gibson (1988: 34) replied: "Perhaps knowledge eventually becomes a system of representations and beliefs about the world (…) but (…) representations and beliefs must be grounded by detection of the surfaces, events, and objects of the layout – the 'stuff' of knowledge must somehow be obtained from the world". With that, Gibson wanted to argue that everything that we perceive, know and experience comes to us through our senses.

Adaptation in ecological theory does not just mean change in the perceiving organism. Adaptation describes the interactive changes in the listener/perceiver and in the environment (Fitch, 2015). As I mentioned earlier in the previous section about the concept of communication between two or more partners, I want to reemphasise the word *resonance* here. In ecological theory, *resonance* is the active engagement of an organism with its environment: shown by actions and spurred by perceptions. Turning toward a sound, or focusing on an object, these actions are caused by perceptions and help sharpen the perceptions in an interactive loop. By interactive loop, I refer to a situation where two or more communication partners are communicating. When information is transmitted, the process of encoding and decoding the available information occurs. Based on mutual understanding, the interlocutors gradually "tune-into" the conversation which begins to flow back-and-forth and that becomes an interactive communication. It creates a perception-action cycle.

The *perception-action* cycle is when, for example, the conversation continues flowing with information. Both participants continue actively listening to one another and engaging by giving feedback. Another example: when two people hike together, they slowly adjust themselves to being in the environment. They see the walking paths, trees and mountains and find their way out by adjusting themselves to the conditions. Then, they photograph what they see and share the photos that were taken with one another. They discover that they both take the same kind of thing. Then they discuss the colours and the shapes of the mountain. In this scenario, the *affordance* of the mountain could be attributed to visual perception such as colours and shapes. And they both share "mutualism [...] and its surroundings, and the idea of affordances as action possibilities of the surrounding that are perceptible by the organism"(Szokolszky et al., 2019: 363).

In a musical context, humans adapt materials to make musical instruments and then adapt themselves to play these instruments. They listen to the music in different ways, such as through live concerts, and on mobile devices. This adaptation is common to the entire human race, or at least of human cultures, but Gibson's theory also describes changes made to individuals through interactions with the environment. The development of musical perceptions can be considered as a sensitisation due to exploratory actions. These learning actions can be directed or indirected. To use ecological theory to determine musical meaning, Clarke (2005) identifies properties of musical sounds that afford certain meanings.

AFFORDANCE AND MEANING

Gibson, who developed ecological theory, coined the term *affordance*, and in his (1979: 127) explanations of the concept, he wrote "the affordance of the environment is what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary. The noun affordance is not. I have made it up. I mean by it something that refers to the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment". The term has since been developed and extended. Gibson was thinking of physical structures like floors, walls and handles. It might now also be used in relation to other aspects of the physical and cognitive environment. For example, some individuals react emotionally when they see red objects. Red, in colour psychology (Pravossoudovitch et al., 2014), is known to signify danger. Red thus affords a reaction to danger or heightened alertness. In design, Norman (1988) also made use of this concept in relation to the comprehensibility of product design objects, specifically buttons and controls and points of contact such as grips and handles: a knob might afford twisting but should also be made to look as if it does. The main point here is that every object affords meaning to the perceiver.

The characteristics (features) of an object (a stimulus) can mean something particular to different perceivers but it is not entirely random. As discussed earlier, slower tempo music affords sadness and the colour yellow affords a happier mood. Researchers also used different musical styles, i.e. the use of atonal music to investigate how their participants associate with it (Mencke et al., 2019). The ecological theory offers an explanation for why listeners often share similar interpretations of music: listeners who have developed their musical sensitivities in the same or similar environments through exposure to a limited repertoire of musical styles are likely to perceive and act upon similar musical affordances (Clarke, et al. 2018).

THE USE OF MUSIC LISTENING AND CONTEXT

We often think of *action* as overt-physical action, such as playing an instrument, dancing and composing (writing the musical notations down on a page). But the act of listening to music is also an action.

Who does not listen to music, especially when music listening has become almost ubiquitous since the widespread prevalence of handy playback devices and smartphones? But why do we listen to music and how do we choose music? Sloboda et al. (2009: 431) summarise it in this way: listeners use music to *distract*, to *entrain*, to *enhance meaning* and to *energise*. The *distraction* reduces boredom; *energising* helps people maintain arousal and focus. Sloboda et al. also offer *entrainment*, which involves performing rhythmic movement along with music, as occurs in dance. *Meaning enhancement* is "where the music draws out and adds to the significance of the task or activity in some way". Listening to music can be used to regulate emotion (Juslin et al., 2001; Van Goethem et al., 2011). Batt-Rawden et al. (2005) emphasised that music catalyses self-extension and self-help. They encouraged participants in a study to activate music by "telling" about music's meaning and its uses in their own lives with others as a way to connect with others and self-empowerment. Music listening can then be seen as an active engagement in a musical event (Curwen, 2020; Clarke 2018). Sounds then can be *for* something as well things to attend to for their own sake. Sounds have associations and we can make associations. Palmer et al. (2013; 2016: 1) explored how other sensory experiences can mutually reinforce effects and found that there are "common emotional associations forming a cross-modal bridge" between music and colour. Bringing the discussion back to cross-modal correspondences, from an ecological theory perspective, these may arise as a result of perceptual learning in a particular context, as people develop sensitivities to relationships that reoccur in their environments over time. What this means in very plain terms is that CMCs are not just plain neurological links between sense systems but have cultural and emotional content too.

In this section, I summarised music perception through the lens of an ecological approach. Gibson's concept of affordance was used in design and in music; the affordance of objects suggests some meaning or usefulness. Based on an individual's capacities, the objects in their immediate environment afford an action. By action, I refer to the less obvious activities such as listening to music. There is previous research observing that people use music listening to achieve a particular purpose as simple as distracting themselves from boredom and but also to energise and motivate themselves. The gist of this is music´s multidimensional aspect beyond the dominant image of the musician and their constrained audiences listening stock still in darkened chambers.

In order to understand the meaning and association of music in individual perceivers, in the next sub-section I will introduce the 4E cognition framework.

2.3.1 4E cognition framework

In a multisensory environment, the question of how affordance can practically apply in designing artefacts has attracted some interest (e.g. Lin et al., 2021; Sutcliffe, 2003). Studies involve touch/motion (Kortum, 2008), even smell (Emsenhuber, 2011; Amores Fernandez, 2020) and taste (Vi et al., 2017). Non-speech audio sounds afford notification when a message is received, but how can a beeping sound or a flashing green light array become more personally meaningful? How can practitioners and researchers use the concept of affordance to design more meaningful sensory experiences to match an individual's subjective condition?

As I discussed earlier in this chapter, most research involves collecting data that tries to study cross-modality by making direct observations of subjects' behaviour (Fitzhugh et al., 2019; Lacey et al., 2016). In this sub-section, I want to briefly address how a 4E cognition framework can fit into researching the more subjective experience of cross-modality; that is to say, the emotionally-mediated correspondence. And how can this framework be used to evaluate the whole communicative system while inquiring about the relations between affordance, artefact, environment and people?

The 4E cognition framework posits that cognition is embodied (based in the body), embedded (in a social-material environment), enactive (actively coupled with the environment in the process of sense-making), and extended (co-constituted by features of the environment) (van der Schyff, et al., 2018; Schiavio, et al., 2020). This framework emphasises the active relationships and co-dependence between organisms and their environment. 4E theorists argue that active musical engagement (e.g., through performing, composing, and listening) emerges as a result of these relationships (Linson, et al., 2017). I will return to this framework in Chapters 4 and 5.

Sense-making is the process of deriving meaning from our experiences in the world. Cross-modal correspondences seem to provide a good demonstration of how sense-making is a function of the individual (and their knowledge and understanding of the world, which includes learned associations) interacting with the stimulus. So the individual brings to the sense-making process abstract knowledge/associations that are not currently present in the environment, but become part of the process anyway.

In summary, in this chapter, I have introduced the concept of crossmodal correspondence and its related methodologies for investigation. I also briefly introduced the concept of synaesthesia. I contend music can be seen as an information and communication medium and that musical engagement is relational and involves other factors (i.e. social-cultural and environmental/natural) in the whole communicative process. Then, I looked into ecological theory such as the idea of affordance and the 4E framework as approaches to study perceivers' subjective listening associations. The theoretical background of this thesis was thus introduced. In the next chapter, I will present my 2-stage empirical study and its methodologies.

CHAPTER 3: METHODOLOGY

This study explored listeners' cross-modal associations in different listening conditions:

1) in a live concert setting and
2) listening to the recorded concert in a non-concert setting (at home or other locations).

Since I was interested in both the emotional and the non-emotional associations that listeners made with the music, I approached the subject with a hybrid, adaptive method. This consisted of experimental and explorative design elements (the visualisation and music performance) and experimental qualitative research methods (open-ended surveys, interviews and observations). This mixed-method was intended to answer my *RQ 0: How is music perception affected by the sensory channel a subject attends to?*

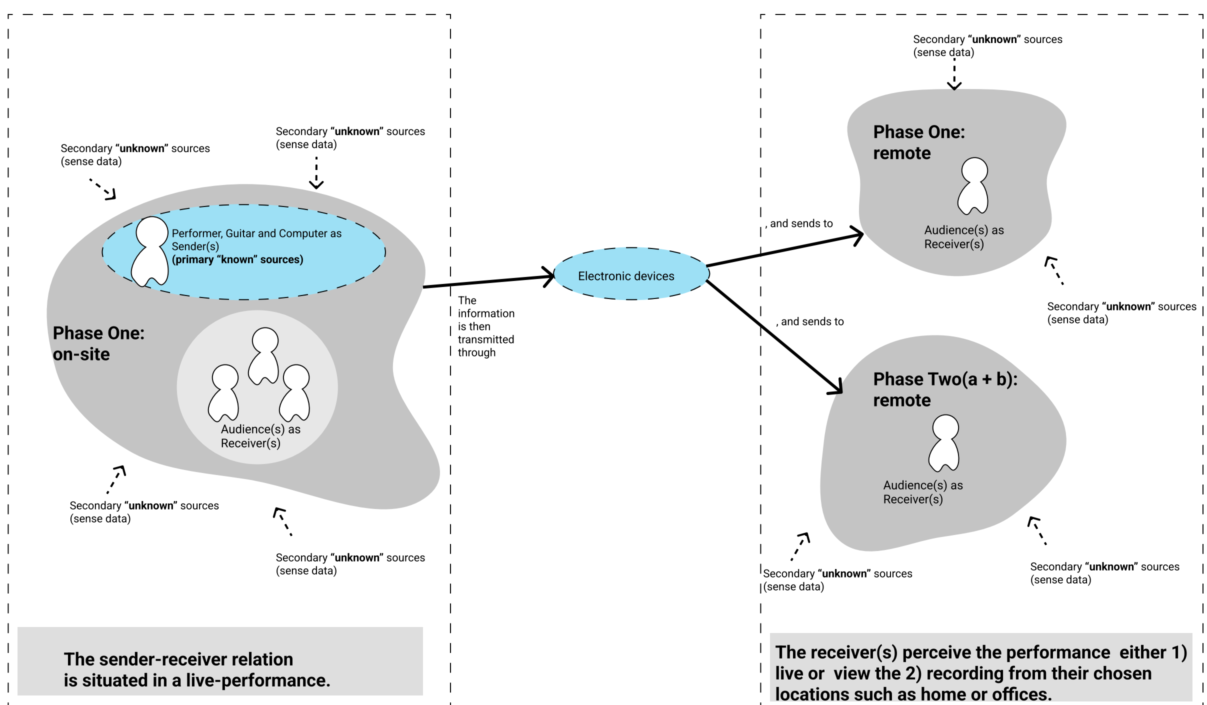


Figure 4. This figure illustrates how the explored theories (in chapter 2) fit into my current experiment.

Furthermore, this 2-stage experiment was designed based on the theoretical background that I have discussed in chapter 2. Figure 4 provides an overview of the relations between *actors* (performer, onsite and remote listeners), *environment* (onsite, remote at home), *artefact* (auditory and visual stimuli) and *social/music interaction* (the interaction between actors and the other elements). Onsite and remote listeners attended a respective concert either live or offline (they listened to the recorded materials). The idea behind this design was to investigate the mediation of a musical communication system and how the different interactions could alter the way listeners understand musical content. Might there be differences between listeners who knew what happened in the concert and who did not know what happened by not seeing how the performance was performed?

An individual´s subjective sensory experience cannot be directly observed, so to get some sense of individual subjective experiences, qualitative research methods such as interviews and open-ended surveys were used. The main question RQ0 and three sub-questions RQ 1,RQ 2 and RQ 3 emerged during the study.

Furthermore, the use of a 2-phase *qualitative* experimental design approach allowed reflections on participants' responses and the validation of the experimental design setups. This approach emphasised the importance of all participants being equally relevant to the research and development process. This experiment drew on the advantages of qualitative inquiry to better understand *meaning construction* and gain a more *holistic view* of responses and differences between the varied grouping of mediated content. Nonetheless, the analysis of the rich and complex reports of participants can be difficult due to the unavoidable translation of first-hand experience into verbal expression. With that caveat in mind, this study used the 4E cognition framework *pragmatically* to interpret data collected via interviews and open-ended surveys.

In the following sub-sections, I will present all the ingredients that made up this qualitative experimental design.

3.1 Experimental design

Phase 1 took place in a concert setting. The concert was held on November 24, 2021 at the Science Library in Oslo. (MusicLab, 2021; MIRAGE, 2022). It was a collaboration with MusicLab, Bjørn Dreyer (the performer) and MIRAGE. Its purpose was to answer RQ 1 and RQ3.

Phase 2 was subdivided into "a" and "b" parts and took place in a non-concert setting where subjects participated from their home or offices in Norway and Ireland. It aimed to answer RQ 2 and RQ3.

3.1.1 Participants

Participation in the study was voluntary. It was supervised by researchers from the RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion (RITMO) after ethical approval was granted by the ethics committee of the University of Oslo. In phase one, the performer (n = 1) was interviewed. 37 participants completed the semi-structured survey (onsite n = 30; remote n = 7).

In phase 2a, five subjects (n = 5) were observed during the use of the app and while watching the video and they were interviewed afterwards. In phase 2b, five participants (n = 5) completed the semi-structured survey but they were not interviewed personally or observed. The general criteria were that participants were active music listeners and creative practitioners who are interested in the concept of music-colour associations.

3.1.2 Procedure

Task:
In phase 1 (see Figure 5 below), participants were audience members for the live concert and heard a warm-up movement that the performer called movement 1 (M0), followed by movement 2 (M1) and movement 3 (M2). In between movements, onsite participants were asked to fill out a printed survey about their immediate impression of the previous movement. For remote audiences, they were asked to fill out an online survey about their experience. The survey contained open-ended questions and Likert scales.

Setup:

The first phase experiment was conducted in an open space on the ground floor of the Science Library in Oslo. It was located directly next to the entrance. During the concert, visitors to the library came in and out frequently. The audience was seated in chairs surrounded by four loudspeakers (commonly known as quadraphonic). The chairs faced the performer, and behind the performer, there stood a screen projector.

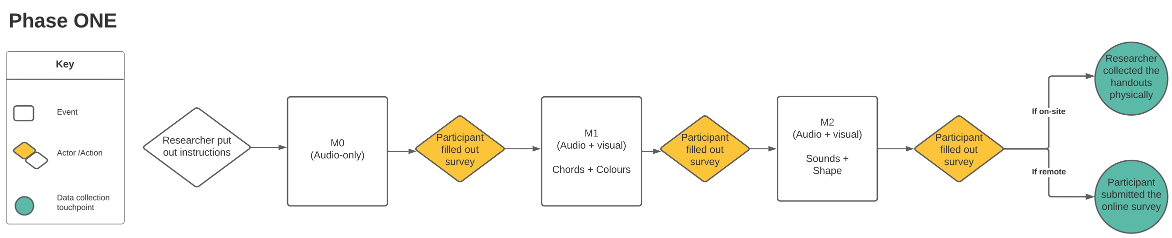


Figure 5. Phase 1 experiment procedure.

Task:

In phases 2a (see Figure 6 below)and 2b (see Figure 7 below), the subjects used an app (Appendix 7) developed specifically for this study, called Synesthesizer (the stimuli were derived from M1). They were asked to assign colours to the music chords. Then, they listened to and watched a visualisation tuned to their colour preferences. Finally, they were asked to view a video of M2 on a Youtube link (Appendix 7). In part 2a, subjects were observed by the researcher and participated in the experiments via Zoom from the period Feb 1 to 13, 2022. In part 2b, participants answered a survey on either a smartphone or a laptop. They watched the visualisation app (M1) and a video of M2 at their own pace and followed instructions. The data collection period in part "b" was from Feb 13 to April 5, 2022.

Setup:

The second phase was a non-concert setting in which 9 out of 10 subjects participated in the experiment from their home, which they reported to be a quiet environment. Only 1 subject participated in a noisy office. Participants were seated in a chair and wore headphones or earphones. They used their laptop for watching the video and a mobile phone or tablet for using the app.

Summary:

While phase 1 was an explorative experiment, phase 2a served as a more elaborative inquiry. Phase 2b was an additional data collection phase that was prepared as a fallback plan.

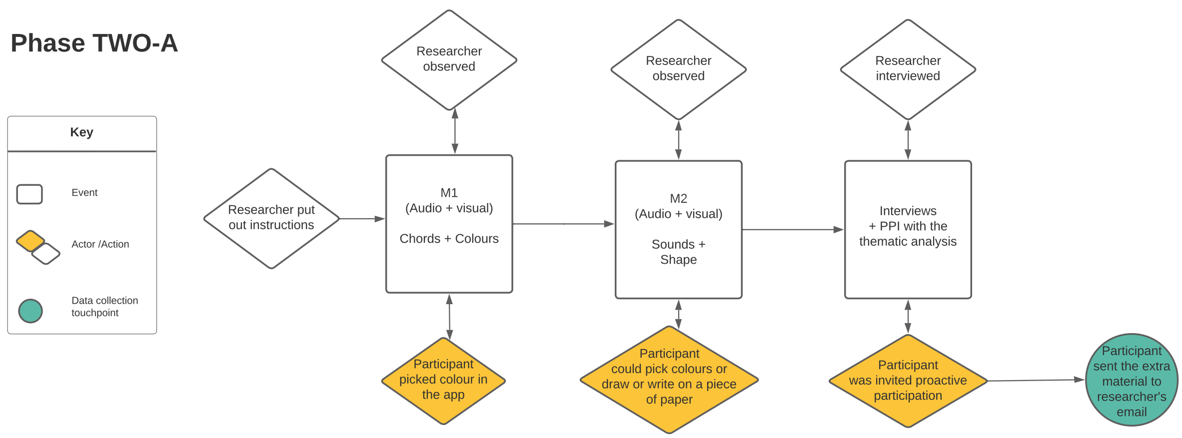


Figure 6. Phase 2a experiment procedure.

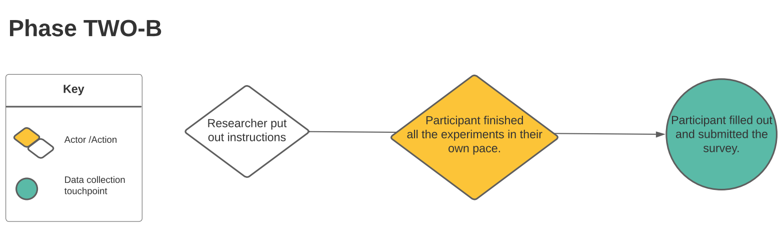


Figure 7. Phase 2b experiment procedure.

3.1.3 Survey questions

In phase 1 and phase 2b, an open-ended survey was handed out to participants. The survey questions were designed to collect first-person's reflections on their own listening experience, such as "*Were you distracted at some point while watching the video?*" The survey items were deliberately generic and unspecific, so as to elicit a wide variety of responses from participants. A similar approach had been used in music cognition research (Schiavio et al., 2020).

In phase 1, it was a live performance in a constructed setting. I was there to observe during the concert, so I did not ask participants about their environment. Instead, I asked about their experience on a Likert scale in order to have a general overview of whether they liked the concert or if there were self-reported synaesthetes amongst audiences. Then, the open-ended questions served to collect the audiences' reflections on each movement. Sample questions such as: "*Feel free to write words or draw something that you associated with the music or visual were asked*." A similar style of questions were asked in phase 1 (Appendix 1) and the aim of asking these questions was to get some sense of individual subjective listening association.

In phase 2b, the survey allowed the listeners to participate in their own time and finish the experiment anywhere they wished. In this context, sample questions (Appendix 1) such as *"Describe your current environment (e.g. in a quiet room, on the street)"* were asked, so as to understand the environment that the participants were in.

3.1.4 Apparatus

In phase one, the concert setting: there were four rectangle-shaped loudspeakers and one large screen projector. The purpose was to create a quadrophonic, immersive, spatial audio space and display real-time visualisation. For online streaming, an open broadcasting software (OBS) was used. It allowed the remote audience to watch the concert live on Youtube. A video mixer was set. It included three cameras from three angles so that the remote audiences could view the live concert from more perspectives: 1) the back of the performer, 2) the close-up of the performer in the front, and 3) the back of the audience. For the visualisation, the video mixer was connected with the screen projector directly, so that the remote audiences could periodically view the correlated visualisation. For the audio experience, an audio mixer was used to connect two microphones in order to capture the music. Due to technical error, the immersive, ambisonic audio experience was not sent to remote audiences, so they only heard the concert with an ordinary audio experience (without special audio effects that were generated by sound engineers during the concert). But this ambisonic effect could be heard from the *recorded* version via headphones.

In phases 2a and 2b, the audiences heard the recorded M1 and M2 on their electronic devices (mobile phones, laptops and tablets, headphones).

3.1.5 Stimuli

In phase one, three pieces of music were played to the audience. The concert started with warm-up music. It was played without visualisation. Participants listened only to the music. Then, visuals and music were played in movement 2 (M1) and movement 3 (M2). The music performance itself was a piece composed by the performer and played on a guitar with live electronics. In phases 2 (a & b), participants listened to and watched the same M1 and M2 from phase one, but the audio and visual information was converted into an interactive app (for M1) and a non-interactive video (for M2). For M1, participants were able to personalise the visualisation based on their own colour preferences while M2 was a video with a spectrogram representation of the music.

Images in Figure 8 (below) show the stimuli in both concert and non-concert settings.

In M1, the visualisation displayed colours depending on the chords played by the performer and tracked by the machine. The colour-chord relations were based on the guitarist's own synaesthetic experience. He was self-reported to have colour-letter synaesthesia (i.e. grapheme-colour synaesthesia). When the guitarist knew which *chord letter* he was playing, the process of thinking about the *"letter"* drew associations with colours. For example, the letter C to him was white and G was dark blue. This movement (M1) was constructed musically with the Western 12-tone scales. The guitarist improvised with diatonic chords on a C major scale. The computer, as a co-performer, created polyphonic continuity by using granulator synthesis. The whole M1 was quiet in volume and slow in tempo. The real-time visualisation displayed a set of colours when it detected guitar chords. For example, when guitar chord A was detected, the colour red (Hex #5A0309) showed on the screen. The computational logic behind this was based on the guitarist' letter-colour synaesthesia e.g. for the guitar chord A was associated with red. Whether it was A major or A minor mattered less than on the letter-colour association. This procedure was followed until the end of the movement.

In M2, the visualisation was programmed to show several layers of the real-time spectrogram of the sound, each associated with a specific colour and placed at a specific depth within a 3D space. The opacity of these layers was controlled in real-time by one operator. The guitarist produced sounds with the use of different sizes of rubber mallets. He rubbed the mallets on either the body or the neck of a blue, metallic guitar. The computer recorded the sound and played it back some seconds later while the performer kept playing and adding more auditory layers.

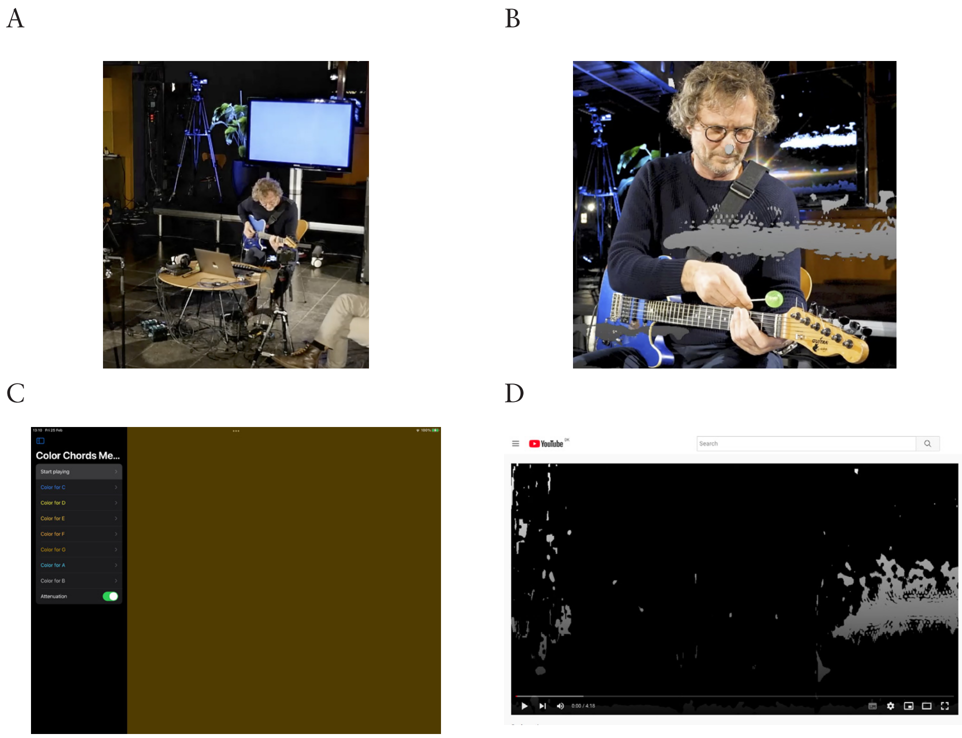


Figure 8. Image A was the setup of M1 in Phase one. Image B illustrated M2 and showed one sound-producing technique that the performer applied. Image C was the Synesthesizer app. It allows users to select colour-chord correspondences and to watch a visualisation of M1. It was used by participants in 2a and 2b. Image D was a remixed video of M2 that was played in phases 2a and 2b.

3.2 Data collection

The data collection is the product of the instrumentation chosen for this investigation (Miles et al, 2017, p. 30). After Kvale and Brinkmann (2015 in Miles et al 2018: 31) "analysis and interpretation occur along the way" so the data collection instrumentation was adjusted from phase to phase. Data from each of the phases were in the following formats:

Phase 1 (Concert-setting)	**Phase 2a (Non-concert setting)**	**Phase 2b (Non-concert setting)**
Audience Survey (paper)		
Audience Survey (online)		Audience survey (online)
	Screenshot of colour choices	HEX colour codes
A/V recording of concert		
Performer's notes	Participants' notes	
Researcher's observation notes	Researcher's observation notes	
	Zoom recording	
Performer interview (Zoom) transcript	Audience interviews (Zoom) transcripts	

Figure 9 – This table (left) is an overview of the data collection methods and stakeholders who were involved in each phase.

CHAPTER 4: DATA ANALYSIS AND EVALUATION

In this chapter the data analysis and evaluation are presented. Content analysis, coding and thematic analysis were carried out. With the ecological theory as a foundation, I interpreted data collected via interviews and open-ended surveys. An additional music performance (video) analysis was conducted.

4.1 Data analysis method

In this section, I explain how the data generated from Phase 1, Phases 2a and 2b were treated. The analysis method originated from Schiavio et al. (2020) and consisted of *immersion*, *categorisation*, and *explanations*. I added an *evaluation* stage after categorisation in order to compare and contrast the data between the "a" and "b" parts of phase 2. This was because there was a slight difference in the data for operational reasons. The steps are graphically represented in Figure 10 (below) with the activities of each step listed e.g. the second step "Categorisation" consisted of thematic analysis.

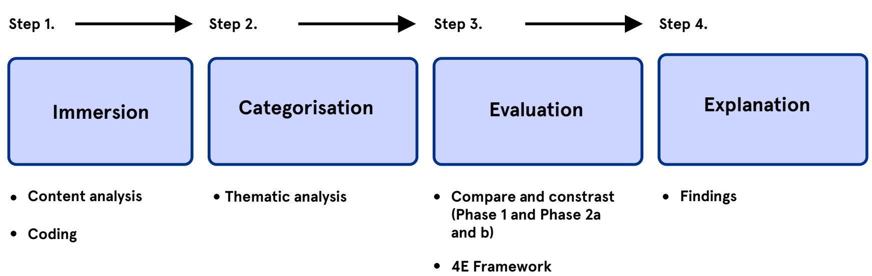


Figure 10. Outline of analysis procedure and tasks involved in each step.

Step 1 involved *immersion*, which was a slow and reflective examination of the material to get orientated in the data. It consisted of content analysis and coding (ideas were tagged). In Step 2, the *categorisation* step, the data from "immersion" was assigned themes (such that tags with commonality are grouped according to relatedness). The *evaluation* in Step 3 involved comparing and contrasting the data from Phase 1 and Phase 2a and 2b. The explanation of Step 4 is dealt with in more detail in section 4.2 below and was of an interpretative/evaluative nature.

For all phases, the qualitative data were analysed inductively based on content analysis (Flick, 2002; Patton, 2014), in which meaningful themes (Braun and Clarke, 2006; Miles et al., 2020) were derived directly from the dataset. In the thematic analysis, the participants from phase 2a were involved in order to check that my categorisations were meaningful. The interviews were transcribed. NVivo software was used for coding. In the coding procedure (Creswell, 2017: 196; Bergin, 2018; Miles et al., ibid; Schiavio et al., 2020), a 4E cognition framework (Van der Schyff et al., 2018; Schiavio et al., 2020) was used pragmatically and adaptively for evaluation. Three types of codes were produced:

1. Emotional codes: these designated the emotional characteristics of the experimental data. Data were plotted on an Arousal scale (calm-excited) and Valence (pleasant-unpleasant) scale (Droit-Volet et al., 2013; Lin et al., 2021; McDonald et al., 2022).
2. Conceptual codes: these presented the main themes in generalising terms of how most participants described each movement (M1 and M2).
3. Colour codes: these generalised the colour choices made by respondents. Data was presented on a Bright/Dark and Cool/Warm colour wheel.

Due to the observation that some variables could not be accommodated by the above-mentioned 3 codes, an additional coding cycle - Attentional codes (Cycle 2) was developed. It produced 2 new themes: 1) Internal attention and 2) external attention. The code sets were integrated into a final set of 4. These were
i) abstract ideas;
ii) memories;
iii) the medium, technical element;
iv) environment/surroundings.

This cycle (figure 11, below) was used to discover some common attentional associations.

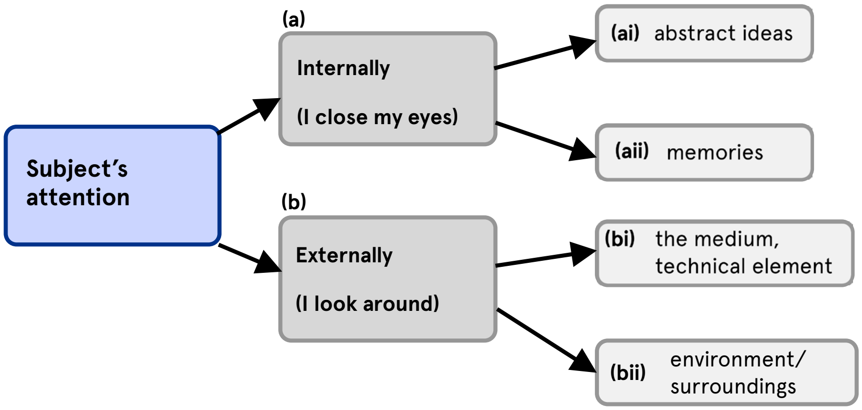


Figure11 (left). The structure of Cycle 2 was based on the observation of codes, and generalisations of the way in which participants described their experiences.

4.2 Coding and findings

This section addresses the key descriptions associated with participants´ responses. These include emotional codes, and colour codes that participants described in response to M1 and M2. In addition, there are thematic categories summarising the multiple conceptual correspondences that participants experienced through mental imagery. Moreover, there are some relevant descriptions of participants' reflections on their emotional or non-emotional associations with the music and visuals.

4.2.1 Emotional codes

For M1, most participants in phases 1 and 2a reported having lower arousal and higher valence. There were exceptions in phase 2b such that some reported having lower valence. The overall emotional code (Figure 12, below) to M1 appeared to be low in arousal, while there were also codes associated with higher arousal and valence. This movement seemed to be emotionally quite positive. Many participants commented on this movement as "*calm*" and "*peaceful*" (Appendix 2,5,6).

In contrast to M2, the emotional association seemed to be unpleasant (lower valence) in all phases. There were some exceptions with a few ratings on the higher valence side in phase 1 and phase 2b. In the live concert setting, participants reported mostly higher arousal while phase 2b (non-concert setting and without researcher's presence) was rated as lower arousal. Audiences commented on this movement as "*mysterious*" and "*uneasy*" (Appendix 2,5,6).

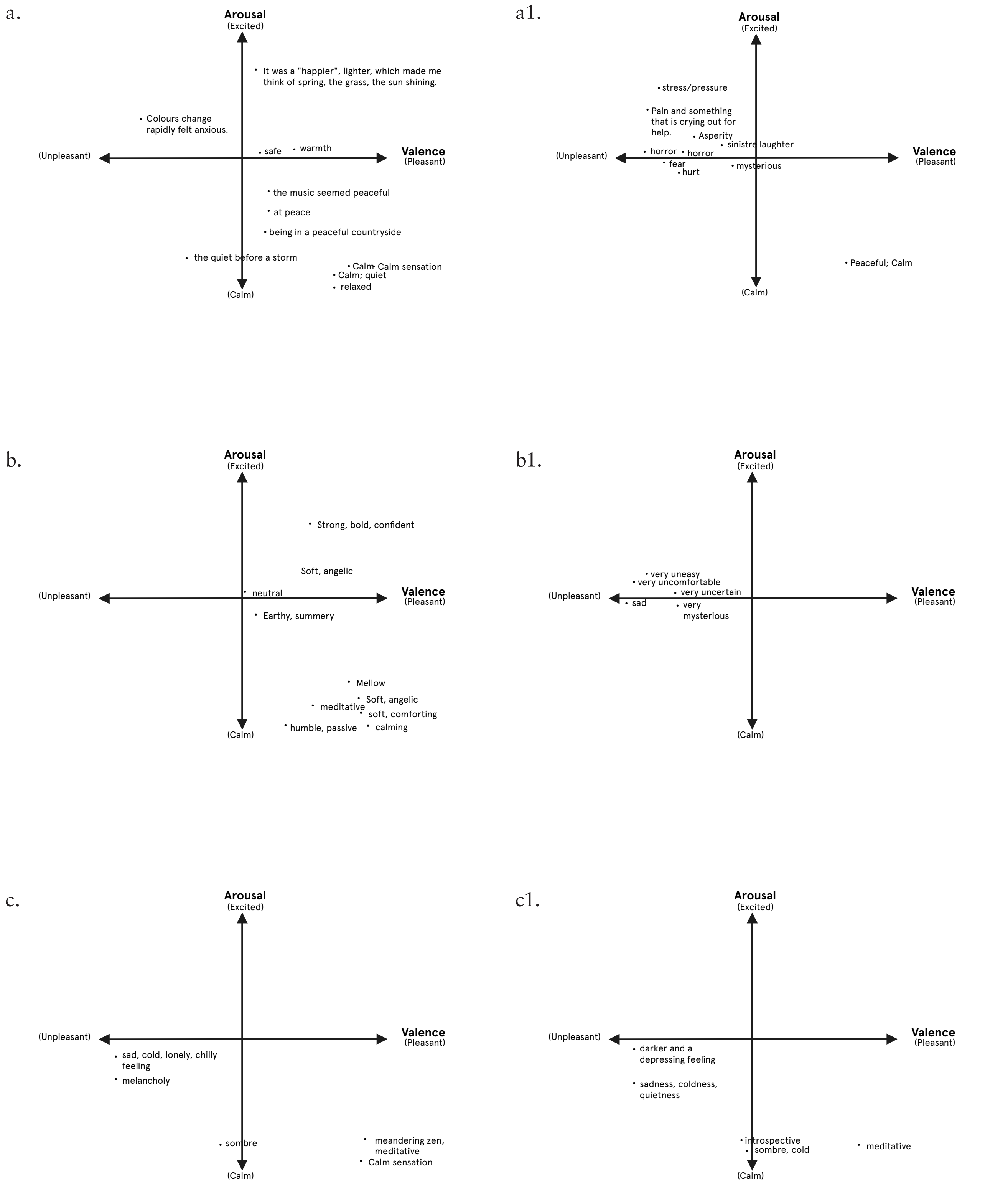
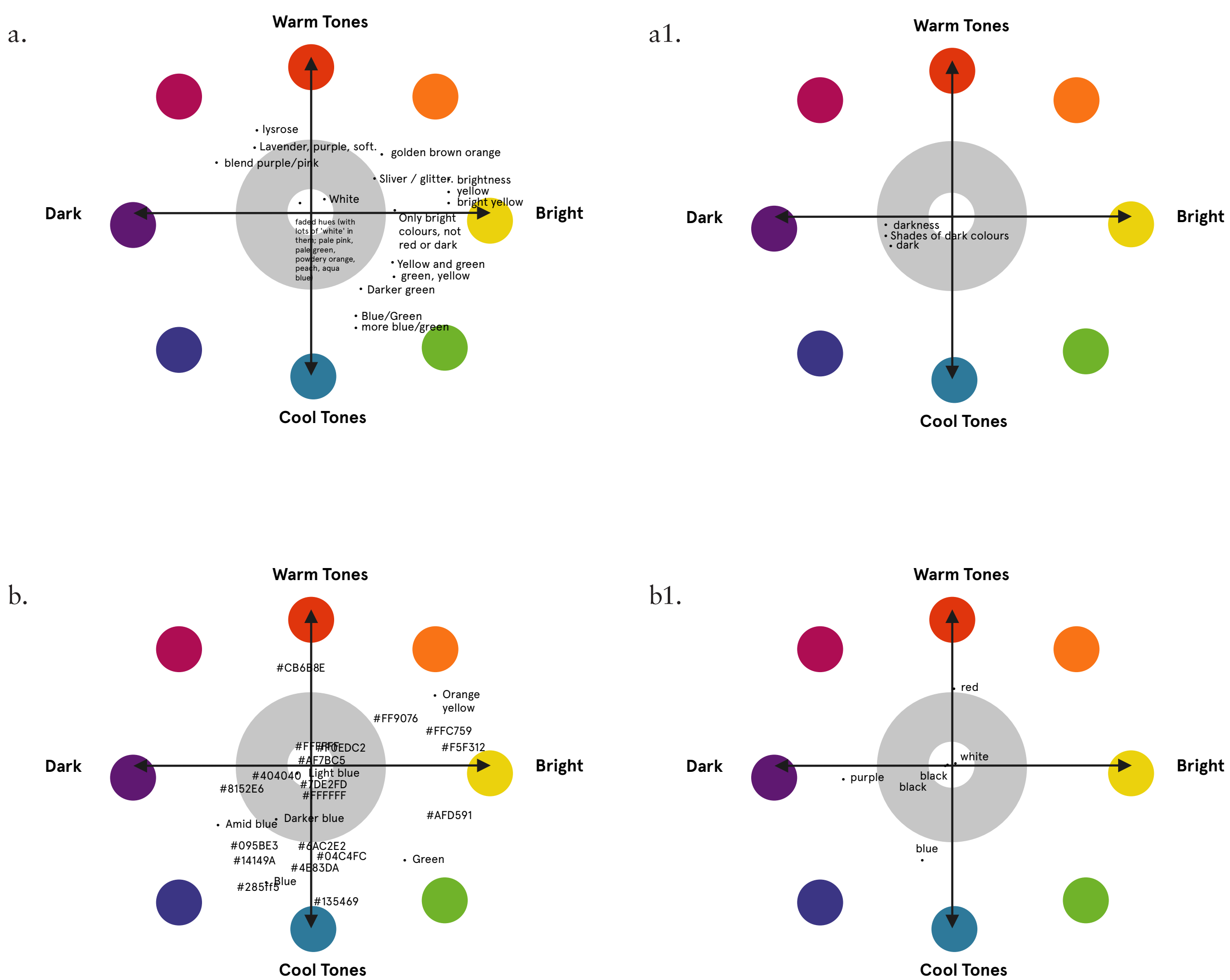


Figure 12. The left column a, b and c shows the emotional codes for M1 in phase 1, phase 2a and phase 2b. On the right column are shown emotional codes for M2 in phase 1 (a1), phase 2a (b1) and phase 2b (c1).

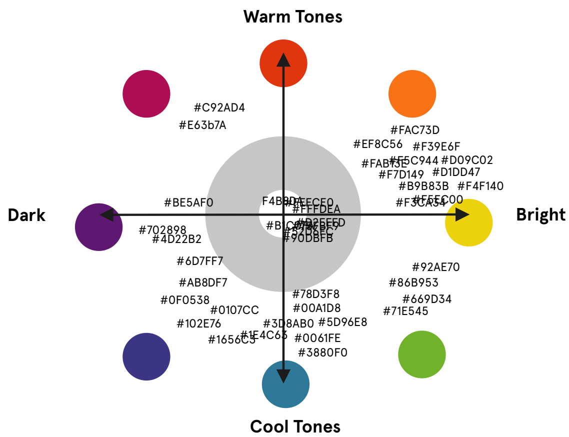
4.2.2 Colours codes

For colour codes, M1 generated more comments about colour than M2. In phase 1, the colour association (Figure 13, below) seemed to be on the brighter side. In phase 2a, the majority of codes seemed to be scattered on both cool and dark (e.g. blue hue) and bright and warm sides (e.g. orange and yellow hue). In phase 2b, the colour codes appeared to have more variables. Inquiries into the motivations that participants had for these colours, some participants in phases 2a and b left comments writing that this movement (M1) was *"green and nature"*, *"calm blue sea"* and *"yellow sunlight"* (Appendix 2,5,6).

M2 received fewer comments on the colour and these tended to be associated with darker and cooler tones.



c.



c1.

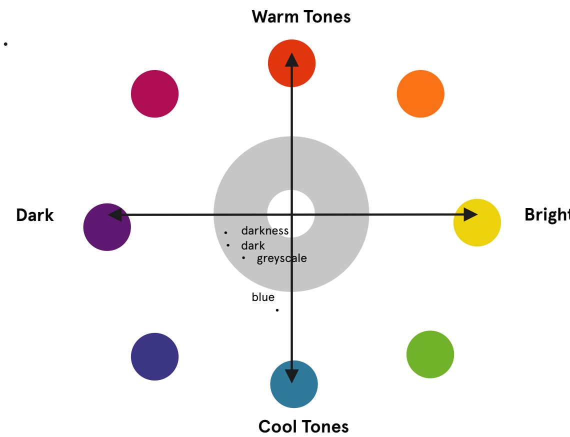


Figure 13. The left column shows colour codes for M1 in phase 1 (a), phase 2a (b) and phase 2b (c). The right column, on the other hand, illustrates M2 in all phases. a1 as phase 1, b1 as phase 2a and c1 as phase 2b.

4.2.3 Conceptual codes

For the conceptual codes (Figure 14, below), the patterns were descriptively micro-coded using NVivo software (which deals with units as small as phrases in a sentence and ones as large as full conversations). I checked how frequently different codes occurred simultaneously so as to identify patterns and themes. A challenge was that "themes" may be abandoned and revisited multiple times within a relatively short span of time. As a result, some codes occasionally overlapped. Categories can belong to several themes in parallel. To reduce the chance of bias, there were 4 participants involved in checking the thematic analysis of all of the data. They agreed or disagreed with my categorisation through the use of a platform called UXTweat. This software gives participants a chance to put terms in categories and so allows checking of the researcher´s choices about the themes.

For M1, all phases shared the same themes. There were themes that appeared in both movements. They were *Nature, Narrative and Emotion*. For M2, both phases 1 and 2 (a and b) shared similar themes. For example, both phases mentioned *Creature and Infrastructure and construction*. Some additional themes arose in Phase 2a and 2b such as Domestic equipment and Folk culture. Participants in all phases commented that the M2's sound was similar to *"whales"*, *"wolves"* and *"crying animals"*. Some notably descriptive comments on M2 were that the sound resembled a *"vacuum cleaner"* or like *"Mongolian throat singing"* (Appendix 4).

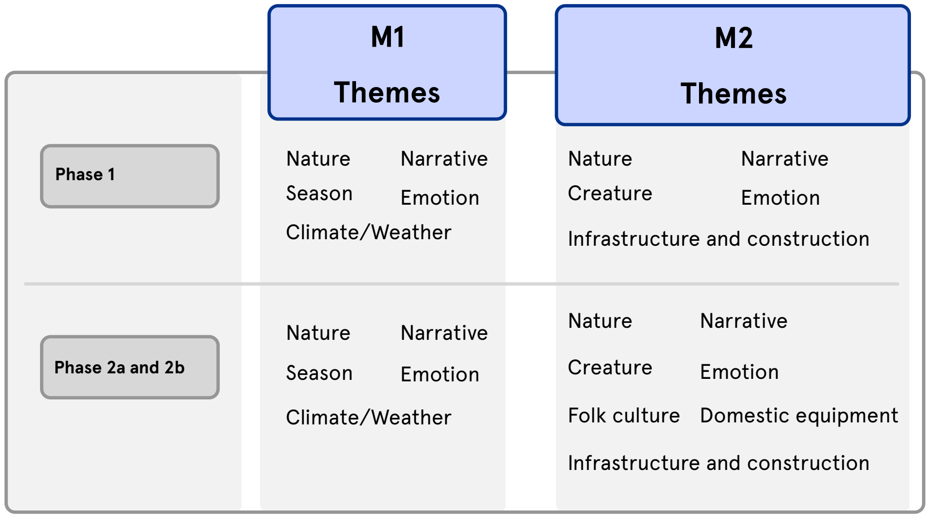


Figure 14. This figure summarises the themes of phases 1 and phases 2a and b.

4.2.4 Attentional codes (2nd cycle coding)

Cycle 2 coding, as mentioned earlier, was developed due to the multiple variables that did not conform adequately to the three codes (emotional, colour and conceptual). This cycle was carried out so as to categorise the factors that participants paid attention to while listening to the music and watching the visuals. This also suggested what could happen when participants decide to close or open their eyes while listening to music. The protocols of interviews are attached to this thesis document under Appendix 2, 5 and 6.

For M1, musically speaking, many participants in all phases mentioned the word *"calming"*. To understand what this meant to the audiences, I looked closer into how participants described their experience, then, I summarised two sub-codes under Internal attention (closed eyes) and external attention (opened eyes).

This decision was also based on my *RQ0: how is music perception affected by the sensory channel that a subject attends to?*

For internal attention, there was evidence that listeners related what they heard to the memories of guitar playing. An example from M1 is this comment *"Was this the electric guitar music? It was very pleasant. I was, a bit too literal, picturing a finger moving over the strings, making a slight squeak" (P03 in phase 2b).* So, music was suggesting images of physical action (which might be visual or muscle-memory impressions). Further examples of these kinds were:

" […] like on the guitar, maybe because I do actually pay[...] attention to what I'm doing, [...] I know A is red, B is yellow. C is blue. G is grey, F is black. D is green, sometimes it's red. If it's D minor is red as well? I think D is grey? D is orange actually. [...] But, I don't know if it's [...] actually the sound. It's probably the shape of [...] my fingers than actually, the sound" .(P05 in phase 2a)

and:

"A sense of past memories, then flashes of fragmented visions". (P05 in phase 2b)

On the other hand, another kind of internal attention could be towards abstract ideas, such as *"I saw a lot of blue […]Like space blue " (P05 in phase 2a)* and *"The first one was more neutral, almost feminine, but not quite more neutral. But the second one was very bold, very masculine" (P01 in phase 2a).* These two descriptions suggested ideas that are hard to articulate or at least ambiguous e.g. space blue. One might ask if that is almost black or is it the colour of the zenith.

Examples of external attention in M1 were where listeners focused their attention outside themselves:

"I was expecting more music and was surprised by how short it was - I would have liked more and more visualisations. It was unfortunate that people were talking throughout the building". (P03 in phase 1)

and:

"The visual was lacking green. I found that a little strange. The music seemed peaceful, and I would think of green as a peaceful colour". (P05 in phase 1)

and:

"I was thinking about: are the colours and chords matching? That's what I was thinking about and I was distracted by that" . (P04 in phase 2a)

These suggested that the participant was aware of their environment and its technical elements during the concert.

For M2, the following selected reflections of participants hinted at the reason why this movement appeared to be on the lower valence side. These quotes deal with attention directed internally:

"it was really [...] the feeling of what's going to happen [was] very uncertain, like very mysterious. but then I'm talking about the music". (P04 in phase 2a)

and:

"I started to think about something completely different […] I started to think about some emails that I was supposed to send. I started to think about the guy from Bangladesh, he is doing some work for me" .(P03 in phase 2a)

These statements coded for attention directed externally:

"[…] what I encountered to hear the combination of two senses. So, in a way, I like the combination to be able to see what I hear. [...] and I can observe in real-time so that I could both see and hear the same thing. It raises my awareness. It added value to my often used listening. So the add-on with the visual part is useful in a way".(P02 in phase 2a)

and:

"So, in the beginning, I was focusing on those shapes, and following them trying to figure out if [...] there's any patterns, any logic [...]behind this. And then, again, after a while, I think my brain was started to [...] fade away, basically, I wasn't following the vital shapes.

I was looking just at one point just registering what is happening around. And then, again, I started to look for some patterns. So, it was like back and forth". (P03 in phase 2a)

The key thing here in these citations is the repetition of the kinds of responses and the way in which they were either internally or externally directed. In the next sections, I analysed the music performance from a video recording that was collected on November 24, 2021, in phase 1 (Appendix 3). The information about the performance techniques and concepts was provided by the performer (BD) in a post-concert interview (Appendix 2).

4.3 Music performance analysis

This section presents a general analysis of the video recorded during the improvised performance in phase 1. BD is the performer. I analysed M1 and M2 only because those were relevant to my research topic about cross-modality. This section is intended to be an impersonal register of the sounds used in the performance. This might give the reader some further insight into the nature of the sounds that triggered the listeners´associations. In the nature of verbal transcriptions of music, it can probably only be indicative.

In M1 (see Appendix 3), from 00:06 to 00:48, BD played the diatonic chords on a C major scale. The chords progressed ascendingly and descendingly between the range from G2 to C5. At00:49, he arpeggiated a perfect fourth as to signify the computer starting to make sounds. It played a constant pitch that served as an accompaniment to the guitar. From 00:57 to 1:01, a guitar cadenza was expressively played. BD used more pressure when plucking the strings. There was also a main chord sequence starting at 1:00. Starting at 2:25, there was an approximately 50-second long transition. The sounds were created by the computer. Granulator synthesis was used to create polyphonic continuity. The guitar joined in again at 3:12, a perfect fifth (A3 – E4) followed by a major sixth (E4 - C5) were played and then a series of interval cycles on high and low registers between the range of C4 to C6, an ascending and descending loop was created. The computer played triads concurrently but on a higher register from the range of C6 to C8. Until 3:42, the guitarist stopped playing. The computer continued playing solo until 4:44. It used the contrasts between registers (high and low) to affect polyphonic continuity. The guitarist played another triad to end on C4 and this lasted until 4:54, the sound faded away. During the 2nd movement, the real-time visualization displayed colours based on the chords detection. Colours changed when the chords changed.

In M2 (Appendix 3), the performance began with the computer making a squeaking sound. At 0:04, the guitarist joined in by rubbing different sizes of rubber mallets (small, medium, large) on a metallic guitar. Sounds were produced. He rubbed the mallets on the body and on the neck of the guitar. The different sizes of the mallets created lower or higher sounds. The real-time visualisation displayed a spectrogram of the detected sounds. The computer was programmed to record the guitarist's produced sounds in real-time. Then, it played the sounds back after approx. 22 seconds. This created a looping effect where the sounds continuously repeated one auditory signal after another. This human-computer improvisation created a duet. The performer himself called this a *"sound art"* in a post-concert interview. His focus in this movement was to continue adding more sounds so that the computer could play more complex textures gradually. BD also called this improvisation a *"choir"*, and he mentioned that he created a *"choir"* by recording the sounds that were produced by rubbing mallets on a guitar. The application of different sizes of mallets created different octaves from low to high that he called *"wave sound"*. At 1:45, the computer produced a new sound but sounded dissonant in some parts. The visualisation changed its colours from black and white to red and green spectrograms were displayed. The dissonant sound changed from quiet to loud frequently. Visualisation from 3:20 returned to grayscale but with less opacity so that it was barely visible. This display lasted until 3:28 when red and blue colours in low opacity returned. In 3:34, the rubbing mallet sounds that were recorded from the beginning were played back. The movement ended at 4:21 when the computer played most of the time, and the performer's main tasks were 1) rubbing the mallets and 2) directing the computer to work in a certain way.

In summary, the M1 was slow, quiet, tonal music. The computer hinted at a sense of modernist music composition, such as the use of granulator synthesis and looping. The visualisation produced the changing colours based on chord detection. M2, on the other hand, featured novel sounds produced by rubber mallets and non-traditional guitar-playing techniques. The computer was programmed to record the sounds in real-time and then played them back. The visualisation displayed a real-time spectrogram when detecting sounds.

In the next sections, I will discuss the main findings of this study within a framework of 4E cognition in relation to music listening conditions and how stakeholders were involved in a musical communication system.

4.4 The 4E framework evaluation

The 4E framework is here used to provide an overview of the *relation* between *actors* (performer, on-site and remote audiences), *objects* (technology, instruments, stationery such as pens and paper, sound-producing tools), and *actions* (writing, improvising, typing, listening, watching, thinking). By inquiring deeper into this relation, I have tried to analyse the possible *affordance* that may be involved in each *touchpoint* during a musical communicative process.

Figure 15 (below) shows an evaluation of this empirical study. Throughout the multiple-stage experiment, I have tested 3 conditions in 3 contexts. In phase 1, which was a concert setting where the conditions were socially constructed in the sense that the roles of performer, listener and general behaviour were quite orthodox. Then, phase 2a was a non-concert setting with my presence to observe and interview the participants via Zoom. Under this circumstance, participants were in a semi-autonomous condition. Then, in the autonomous condition in phase 2b participants chose to participate in the experiment in their own time.

Referring to Figure 15, I would like to explain the way in which the contents were assigned to the 4E categories.

-In the *embodied* row, all the verbs were highlighted in red and *italic*. Cognition is mediated by the person´s behaviour.

-In the *embedded* row, the red highlights in *italic* nouns of the technology, and how actors interact with the music performance.

-In the *enacted* row, I have highlighted the actions in red and *italic*. It indicated the actions of actors. It categorises what actions the actors were taking in the immediate environment they were situated in.

-In the *extended* row, I listed the objects that actors come in contact with when they interact with the music performance.

4E (1)	Phase 1			Phase 2a		Phase 2b
	Socially constructed (Concert setting)			**Semi-autonomy (Non-concert setting)**		**Autonomy (non-concert setting)**
embodied	Performer *plays/improvises, listens, sits, thinks, feels*	Onsite Audience *writes/draws, listens, watches, sits, thinks, feels*	Remote Audience *writes/draws, listens, watches, sits, thinks, feels*	Audience *listens, watches, writes/draws, sits, thinks, feels, talks, sings*	Researcher *types* notes while observing.	Audience *picks* colour in the app, *listens, watches, thinks, feels*, and *types* in the survey.
embedded	Performer improvises with a *computer* and the sound plays back from the *loudspeakers*.	Onsite Audience listens and watches the concert *live* and through the *audio/visual array*.	Remote Audience listens and watches the concert *live* with personal *electronic devices*.	Audience listens and watches the materials in personal *electronic devices* while turning *Zoom* on.	Researcher observes the audiences via *Zoom* on her *laptop*.	Audience listens and watches the materials in personal *electronic devices*.
enacted	Performer *improvises* in a concert room.	Onsite Audience *sits* in a concert room.	Remote Audience *sits* at home.	Audience *sits* at homes/ in the offices.	Researcher *sits* at her own home.	Audience *is situated* in chosen environment.
extended	Performer: Computer; Guitar; Rubber mallets	Onsite Audience: Pen and paper; Computer	Remote Audience: (1)Laptop; (2)Headphones	Audience: (1)Pen and paper; (2)Laptop; (3)Mobile devices; (4)Headphones	Researcher, (1)Laptop,; (2)Headphones	Audience: (1)Computer; (2)Mobile devices; (3)Headphones

Figure 15. This shows the physical/material/environmental, directly observable factors of the musical communication systems in this study.

Figure 15 shows the *physical* aspect of musical interaction and the engagement of actors. However, I would contend that this is the first level of using a 4E cognition framework, showing the relations within a musical communicative system. It shows what is there but does not explain it. I would like to introduce a second level of using this framework, in which I would focus on the *cognitive/affective* aspect of individual actors. This deals with why the participants described what they did.

In the boxed sections, below, I offer examples of how to frame the 4E. This is based on examples and is intended to aid comprehension of the *cognitive/affect* aspect of the music listening experience.

***Embodied* Listening: Mind in Body**

Music listening in 4E theory is tied to the emotions we feel in our bodies. People can choose a colour based on the emotional content in the music as they have felt it. An example in phase 2a, some participants chose to sing back (emotion is voiced with the larynx and lungs) the melodies and picked the colours (a visual phenomenon), some preferred to play the music chords with their own instruments (fingers, arms) so that they could "feel" the colours and assigned it to a particular chord. In other phrases, some participants expressed that they were trying to figure things out. Active thinking or trying to make sense of the immediately available information requires the engagement of the whole body. The body is central to human cognition and information-processing. Active listening requires the perceiver's mental and perhaps emotional engagement in order to make sense of the music. In this aspect of 4E, the listening is not just in the head, via the ears. It seems to travel back out to the hands, lungs and eyes.

Embedded **Listening: Mind in Culture**

Humans are *encultured* entities. We have learned to adapt to culture and its associated technology. The act of listening to music activates our emotional and cognitive connection with our prior knowledge and experience of culture (which means social-specific values, meanings, customs and beliefs). The listening experience is embedded not only in the technical systems but also in our cognitive/affective systems. It includes our social understanding of something. It is the culture(s) that we embedded through different stages of life, as lived experience and knowledge. A sound brings back memories that are personal to us but also memories that are collective e.g. knowing the meaning of the national anthem or a famous pop song.

An interesting example is found in phases 2a and 2b. Participants thought of their own guitar playing when listening to the sounds. Another example to do with the associations that listeners made with certain sense stimuli: In M2, there were two major words that participants from all phases mentioned, they were *"wolves"* and *"whales"*. Since my participants were based in Norway and Ireland, the choice of an animal might be influenced by the Norsk or Celtic mythology. In Ireland, wolves were commonly known as mythological creatures. In Phase 2b participants noted imagery to do with damp/flooded underground spaces. As well as being personal imagery, this might be associated with the kind of incidental music used to dramatise such scenery in film and television drama.

Enacted Listening: Mind in Action

Our knowledge and experiences shape our decisions and actions. In a multisensory environment, what we pay attention to reflects our own lived experience. The way we respond to our immediate environment comes from our capacities. In the empirical study, professional musicians and non-musicians participated in phase 2a. When a professional musician was asked to pick a colour for the chords, immediately, she picked up the guitar that was near her. Then, she played the chords by herself and she imagined colours while playing the chords. She said that she was *"actually paying attention to [...]the shape of [her]fingers."* Then, she thought of colour for a particular chord. The non-musician who expressed this did something more appropriate for their experience and capacity: *"I was trying to relate chords with colours to check if my selection makes sense, but it looks a bit random"*. This finding suggested that experienced musicians may engage in active listening and thinking of the music in parallel with their own actions while the non-musician appeared to separate the listening experience from their own experience of playing the music.

Extended Listening: Mind in (Virtual) Environment

The concept of *extended* has been commonly explored from the more obvious perspective of the music performance (Van der Schyff et al., 2018) and composition (Schiavo et al., 2020). What happens when this concept is also applied to the music *listening* experience? Van der Schyff et al. (2018) write, "the extended dimension of the 4E framework holds that other features of the environment can co-constitute the mind " so the non-biological features could be seen as musical instruments or composition software in the above-mentioned examples. From the performer´s perspective, he is both playing (performing) but also listening to the feedback from the computer. The computer and the speakers are the performer´s environment. For the listener, it is less clear how the listening might be seen to be extended. During the phase 2a experiment, one of the participants was situated in a noisy office. Instead of not being disturbed by the noise in his immediate environment, he was focusing on the experiment *"because of my [his] earphones"*. Then, he added: *"I actually focused on the music because you [the researcher] asked me to, well, obviously the earphones helped a lot."* In this context, the extended dimension of listening experience could be involved with the (1) earphones, to cancel out the immediate noisy environment and (2) other minds, in this context, the researcher on Zoom. It is at best a weakly demonstrated instance of extension. What this could look like in an ideal case might be other individuals listening at the same time, making sense of the music and reacting to one another e.g. a crowd responding to music and this collective response affecting the individual.

In summary to this section, I have explored how the 4E framework could be used to analyse music listening experience from 1) *physical* and 2) *cognitive/affective* aspects. More ongoing testing with this framework is required though. However, in my current study, I gave an example from one of the many angles in order to check how this framework could be used in this context.

In chapter 4 as a whole, I have presented my interpretations and findings of phases 1, 2a and 2b. I have produced 4 types of codes 1) emotional, 2) colour, 3) conceptual and 4) attentional. In the analysis, there were similar and contrasting patterns between phases. The music performance analysis provided some objective knowledge in order to understand how the musical structures may affect listeners' perceptions. For example, M1 was tonal music in C major with a higher range up to C8. M2, on the other hand, was a novel sound created by the rubber mallets. The dynamic in M2 was both quiet and loud while M1 was mainly quiet. From the data, participants indicated lower arousal and higher valence in M1 while the M2 responses were mostly lower valence and darker colours. How a mode of music performance was played seemed to affect the listeners' associations. In addition, the 4E framework provided an overview of the empirical study and all the relations between actors, actions, affordance and artefacts. I evaluated the experiment from both *physical* and *cognitive/affective* aspects in order to check to what extent can a 4E framework make use of investigating a music communication system. In order to draw a conclusion to these findings, I will discuss some key findings in chapter 5 and provide a conclusion for my current thesis.

CHAPTER 5: DISCUSSION AND CONCLUSION

In Chapter 5 I conclude and discuss the findings of the research questions. In section 5.1, I open some general discussions arising from this empirical study. In section 5.2, I want to specifically answer my research questions and propose future theoretical considerations and open for further discussions. Finally, I will provide a final comment on this thesis in section 5.3.

5.1 General discussions

The qualitative study was designed to explore the complex scope of listeners' cross-modal correspondences. It investigated whether or not there were common experiences that listeners shared when they are situated in different listening conditions (live performance and home listening). Based on my findings, the participants provided some key insights which emphasised the way in which music listening 1) evokes emotions and 2) elicits mental imagery. In the following, I address a few key points that are open for discussion.

COLOUR-MUSIC ASSOCIATION AND CERTAIN EMOTIONAL CODES SEEM TO BE RELATED

In this project, music listeners seemed to relate to certain emotional codes with colour-music association. In this case, emotions might mediate the music-colour associations. This general finding aligns with much previous research in music and emotions e.g Palmer et al.´s (2016:1) studies. They found "common emotional associations forming a cross-modal bridge between music and colour" (ibid: 1). In my current study, my participants seemed to share similar findings as Palmer et al. For example, my *colour* and *emotional codes* suggested that the sound of the rubber mallet dragging on a metallic object links to a darker colour palette, higher in arousal but lower in valence emotion. Meanwhile, my exploration (M1) suggested that a slow pace and quiet tonal music go with the colours blue, green and yellow, which are cooler tones. Many participants found it was "*calming*". An explanation for this could be the timbre of the guitar; harps and guitars have traditionally been thought to have a healing effect (Thaut, 2015).

For future research, I would like to conduct a correlation analysis (Lindborg et al., 2015) in order to examine further the correspondence between music and colour. The reason I could not conduct a mixed-method analysis this time was that it was outside the scope of this study. For the hex colour codes that I have collected from phases 2a and 2b, I could analyse the colour space that is correlated with the music. One could also move beyond simple, single colours and see if paired colours produce a different effect or if musical stimuli could be mapped to more than one colour. More difficult but interesting might be trying to test what music response would a listener associate with an emotional or colour prompt.

MUSICAL IMAGERY

Spence (2013) discussed cross-modal mental imagery. The imagined responses that people make to music can either be triggered internally in the absence of any clear external inducer (e.g. when you spontaneously want to recall the appearance of your childhood home) or can be associated with the presentation of a sensory trigger (as when the timbre of a fiddle suddenly summons impressions of an event, the people present and one´s place in the setting). As I discussed earlier, music is an embodied activity. It is integrated into a perceiver‘s lived, emotional experience. A piece of music affords listeners to imagine the body movements involved in playing an instrument, as well as emotions and images. In my current study, all the generated codes (see: Chapter 4) suggested that listeners shared some kind of mental images. In my conceptual codes, participants seemed to find associations with the themes of *Nature, Emotions* and *Narrative* for both M1 and M2. The mental imagery represented listeners' lived experience. In this context, it appears that the use of open-ended surveys and semi-structured interviews is suitable to study the cognitive aspects of music listeners and their cross-modal associations.

ATTENTION AND DISTRACTION

Another key finding involved distraction. There were two types of distractions discovered in this study. The first type came from the external environment, for example, participants in phase one were distracted by the *"people [who] were talking throughout the building"* and requested that they *"need a quieter environment"*. Another type came from the initial external trigger, that the participants initially perceived the technical elements, and then, felt distracted. Examples such as *"[...]Maybe because of the music, which is a bit slow. I felt for a moment like going to sleep. [...] And then, when I was looking on the phone, at some point, my vision [got] blurry, I just lost focus. Because it was like [...] look[ing] into the phone* *and the colour changed, so, [...] maybe my brain…I just lost my focus" (P04 in phase 2a).* An interesting finding related to a participant in phase 2a who was situated in a noisy office during the experiment. Unlike phase 1 where some participants were distracted by their immediate noisy environment, the participant said *"No [not distracted by the noisy environment he was in], not really, because of [his] headphones" (P04 in phase 2a).* In this case, the participant took action to limit the extent to which their listening experience was embedded (and subject to shaping by) their immediate physical environment. Perhaps one could inquire into how visuals for music performance aid or detract from concentration. Often music is paired with visual material that is supposed to add to the experience. Perhaps some attempts to do so do not succeed as planned because the visuals do not cross-modally correspond. We could inquire further into minimum and maximum levels of secondary stimuli that aid or detract from a musical experience. Or, we might consider which sound structures support a visual signal.

THE 4E FRAMEWORK FOCUSES ON THE RELATIONSHIPS IN MUSICAL COMMUNICATIVE SYSTEMS

From a 4E perspective, the experiment started from a composer-performer's creative generation of music. I would describe his performance as enactive, as it was subject to real-time shaping based on feedback from the computer. The performance was then embedded in the concert setting. It was also recorded for offline listening, offering the possibility of future listening experiences that would be embedded in other environments. For the audiences who comprised my participants, the music, visualisation and the physical devices that were displayed to them became, in some ways, an extended part of their cognitive processing system. During the experiment, participants listened to and watched the music and visualisation and based on my findings, some were actively thinking and imagining. Some referred to the music of their own musical practice, *i.e.* imagining their embodied finger movements. Some associated the sound with emotions, imagining that their feelings linked to being out in nature. Despite the fact that the music listening experience may seem passive at first, my current findings and much previous research found that music is an active engagement. It affords not only overt-physical action like dancing but also inward-orientated thinking and imagining. The 4E cognition framework opens up interesting possibilities to reconsider the generative role of musical engagements. With this in mind, it may provide practitioners with a more solid foundation when designing cross-modal musical events (i.e. staging, sound array positioning, visualisation, and audiences' participation). The people (performers, audiences, remote listeners) who are involved in musical communication systems are all relevant. If technology does not consider the coordination between *actors* (audience, performer), *environment, artefacts* (multisensory), and *social interaction*, it may fail to convey an effective message.

Based on this study, the fourth E, *extended*, seemed like the one that proved the hardest to find a basis for. It could be that my study was not designed to test for it (e.g. maybe I did not ask the "right" questions to find out in what ways participants' cognition was extended). Or perhaps music listening is simply not suited to this except under highly artificial circumstances. But, perhaps those artificial circumstances might be as revealing as visual illusions are for visual cognitive psychology (Klopfer, 1991).

PERSONALITY TRAITS AND THE FAMILIARITY OF MUSIC STIMULUS

I end this section by discussing possible future work. In my findings, there were more participants who liked M1 than liked M2 (Appendix 2, 5, 6). Some participants said that the novel sound in M2 was an *"awkward and rather an unpersuasive mode of musical performance"* (P10 in phase 1). Although I did not know the exact reason why, some additional research on personality traits, such as the Big Five personality traits (Boccia, 2020) may be useful to understand what personality type can appear to enjoy unfamiliar, novel auditory and visual stimuli. If HCI demands more use of sound, it would be good to know the general preferences of subsets of users rather than treating them as a homogenous group.

Finally, familiarity versus novelty. I found that the subjects responded ambiguously regarding the familiarity of the sound. Some participants appeared to wonder what the sound, i.e. M2 was:

> *"I liked the music, and the treated voices (?) at the start" (P03 in phase 2b).*

and:

> *"[...]because of the long sounds like long soundwaves [...,] I had the feeling I was standing in [...] mist, like a big fog and I couldn't see [?] around me [but] only myself. [...] then the images just got very chaotic. [...]but then, again, matched with what I heard. [...]. [...] it was just a lot at one time, [...], so I had to listen while watching a lot of images. So,[...] I got*

a lot of [...] stimuli?" (P04 in phase 2a).

This is in line with work by Omigie et al. (2022) and Mencke et al. (2019) that indicates that care needs to be taken with the choice of sound. Sounds used to attract attention may need a different character than sounds used for other purposes. What sorts of sounds are intelligibly familiar but not so familiar that they fail to attract attention? The findings suggested that the level of familiarity may affect the way in which subjects feel and think about the auditory stimulus. Is this a culturally specific matter or are there cross-cultural tendencies? Work conducted on a sound that was adequately novel to gain attention might not be valid in other cultural groups.

5.2 Specific discussions

This section is intended to specifically answer my research questions in this project. Despite the fact that there are always other angles and other perspectives to answer these questions, I will try to respond based on the findings of this work.

RQ 0: How is music perception affected by the sensory channels a subject attends to?

In the present study, I discovered two major listening behaviours based on the participants' responses: 1) with eyes closed and 2) with eyes opened. Some participants preferred to integrate hearing and vision while others prioritised hearing and suppressed vision. Shutting off the vision helped some listeners to focus on the music but perhaps this is a response to new sounds. We might also have confronted listeners with familiar music and noted their wish to see or not see while listening. From my data, I learned that most participants relate to music based on their own memories and abstract ideas about the music and that these imagined associations are cross-modal (e.g. experiences of nature). There were also participants who observed the technical elements of stimuli. Some participants' attention landed on the auditory and visual features of their immediate environment, too. Furthermore, some participants expressed that they prefer seeing certain colours for the visualisation because the music made them feel a certain way. RQ0 remains an ongoing discussion.

RQ 1: What is the relation of the listening conditions to the way the music content is understood in a concert setting?

In Phase One, in a concert setting, participants seemed to pay attention more to the content in an analytical way and focused on the mental representation of the music. They were able to see how the sounds were created *live*. Being able to see the correlation of music and the performing techniques may have helped listeners understand the content. They were at times disturbed by the noisy environment when the music was not loud. One possible reason to explain this was that no one was able to wear headphones during the concert.

RQ 2: What is the relation of the listening conditions to the way the music content is understood in a non-concert setting?

In Phases 2a and 2b, people seem to pay attention to the feeling of the sound. One possible explanation for this is that participants did not know how the sound was made. It appeared that more subjects focused on the emotional content. However, most reported experiencing more mental imagery in the 3rd movement (M2) due to the *unknown* sound. They did not mention any disturbance by the environment they were situated in. But, rather, they reported being disturbed by the *"unpredictable"* and *"abstract"* kinds of stimuli. Because of this, one of the participants reported *"los[ing] focus"* and *"starting thinking about work email"* while most started *"thinking/figuring what was it all about"*. They wore headphones during the listening experience, so most were reported to be *"very concentrated and focused"* during the experiment. Being self-critical, the research design ought to have stipulated the avoidance of headphones. In a way, the headphones constrain or de-extend cognition, isolating the participant from their environment.

RQ 3: What is the difference between these two sets of results (RQ1) and (RQ2)?

The two sets of results suggested that different listening conditions alter the effect of the music on the listeners: the concert setting (Phase 1) led to responses with higher arousal than in phases 2a and 2b. Concert participants seemed to have more salient signals coming at them during the performances, so it would have been harder for them to adequately attend to everything. Although there may be no single explanation for this pattern, I could refer to the music performance analysis that M2 was a novel sound produced by rubber mallets rubbing on a metallic guitar. The sound was dissonant and was periodically quiet and loud. Compared to M1, M2 had more dynamics. A reason why participants in the concert setting reported high arousal could be that they heard the dynamic of the sound in real-time and onsite. It could be the recording had softened the sound or listeners in a non-concert setting did not listen at a high enough volume. A simple factor might be the psychophysical effect (vibrations or sound waves felt in the chest cavity, for example) of a powerful four-speaker set-up compared to the smaller speakers available to the non-concert listeners. An additional reason could have been the social setting: the fact that they were experiencing the music together with other people.

5.3 Conclusion

This project explored music listeners' cross-modal associations. A 2-stage experiment was conducted in a live concert setting and an offline non-concert setting. A ecological cognitive approach was applied as a foundation of the project. Data collection methods such as interviews, observations and open-ended surveys were used. Due to the complexity of participants' responses, I used content analysis and coding to categorise the variables into themes and tags. Four major codes were generated: *1) emotional, 2) colour, 3)conceptual and 4) attentional*. These provide future consideration of the different factors that cross-modal music listening research can be further looked into. A mixed-methods analysis is recommended *for future research* such as the correlation analysis to investigate individual feature-to-feature correspondence. Moreover, I found that timbre (i.e. the sound of a guitar or plucking string instrument) affords a calm sensation and elicits the cooler tone colour (i.e. blue, yellow and green). Also, the use of a C major, quiet in dynamic and slow in tempo appeared to elicit imagination to the themes such as *Nature* and *Narrative* such as walking in nature or a forest. This cross-modal mental imagery effect of music suggested that listeners could use slow, quiet, guitar/plucking strings music in combination with cool tone colours as an emotional regulation strategy. In HCI, this could further be tested and implemented as a music therapeutic tool. Another finding between music and imagined finger movements and colour suggested that using colour and music or even letters to teach motor skills to young children may be feasible.

The general conclusion I draw from the findings of this work suggests that cross-modal correspondence is a potentially rich topic of exploration that the small-scale nature of this project only touched upon. The relation of music to vision and emotion is a three-way relation of three complex phenomena with numerous variables. Part of the phenomena are based on physical causal relations and part of them are inevitably cultural and related not to causation but to meaning and interpretation. This obvious conclusion is that research in music requires all the tools available to researchers to adequately do justice to music´s physical, cultural and emotional aspects.

References

1. Adeli, M., Rouat, J., & Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. Front. Hum. Neurosci., 8, 352.
2. Amores Fernandez, J. (2020). Olfactory interfaces: toward implicit human-computer interaction across the consciousness continuum (Doctoral dissertation, Massachusetts Institute of Technology).
3. Barbiere, J. M., Vidal, A., & Zellner, D. A. (2007). The color of music: Correspondence through emotion. Empirical studies of the arts, 25(2), 193-208.
4. Baron-Cohen, S. (1996). Is there a normal phase of synaesthesia in development. Psyche, 2(27), 223-228.
5. Batt-Rawden, K., & DeNora, T. (2005). Music and informal learning in everyday life. Music Education Research, 7(3), 289-304.
6. Bergin, T. (2018). An introduction to data analysis: Quantitative, qualitative and mixed methods. Sage.
7. Bishop, L., & Goebl, W. (2018). Performers and an active audience: Movement in music production and perception. Jahrbuch Musikpsychologie, 28, 1-17.
8. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.
9. Bramwell-Dicks, A., Petrie, H., Edwards, A. D.,

& Power, C. (2013). Affective musical interaction: Influencing users' behaviour and experiences with music. In Music and Human-Computer Interaction (pp. 67-83). Springer, London.

10. Brattico, E., & Varankaitė, U. (2019). Aesthetic empowerment through music. Musicae Scientiae, 23(3), 285-303.
11. Bresin, R. (2005). "What is the color of that music performance?," in Proceedings of the International Computer Music Conference - ICME 2005, (San Francisco, CA: International Computer Music Association), 367–370.
12. Burke, J. (1978). Connections. Macmillan London Limited.
13. Boccia, I. (2020). Associations between Personality Traits and Music Preference.
14. Cespedes-Guevara, J., & Eerola, T. (2018). Music communicates affects, not basic emotions–A constructionist account of attribution of emotional meanings to music. Frontiers in psychology, 9, 215.
15. Clarke, E. F. (2005). Ways of listening: An ecological approach to the perception of musical meaning. OUP, USA.
16. Clarke, E., Williams, W. A. E., & Reynolds, D. (2018). Musical events and perceptual ecologies. The Senses and Society, 13(3), 264-281.
17. Cook, P. R. (Ed.). (2001). Music, cognition, and computerized sound: an introduction to psychoacoustics. MIT press.
18. Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.
19. Cross, I., & Woodruff, G.E. (2009). Music as a communicative medium. The prehistory of language (Vol. 1, pp. 113-144). Oxford University Press, 2009. Oxford.
20. Curwen, C. (2020). Music-colour synaesthesia: a sensorimotor account. Music Scientiae. SAGE.
21. Cytowic, R. E., & Eagleman, D. M. (2011). Wednesday is indigo blue: Discovering the brain of synesthesia. MIT Press.
22. Cytowic, R. E. (1989). Synesthesia: A union of the senses. Springer. New York.
23. Darrow, A. A. (1990). The role of hearing in understanding music. Music Educators Journal, 77(4), 24-27.
24. Deroy, O., & Spence, C. (2016). Crossmodal correspondences: four challenges. Multisensory Research 29, p. 29-48. Brill.
25. Dreksler, N., & Spence, C. (2019). A critical analysis of colour-shape correspondences: examining the replicability of colour-shape associations. I-Perception, Vol. 10(2), 1- 34. SAGE.
26. Droit-Volet, S., Bueno, L. J., & Bigand, E. (2013). Music, emotion, and time perception: the influence of subjective emotional valence and arousal?. Frontiers in Psychology, 4, 417.
27. Emsenhuber, B. (2011). Scent marketing: Making olfactory advertising pervasive. In Pervasive advertising (pp. 343-360). Springer, London
28. Espeland, K.M. (2006). 'Color for the eye, color for the ear': music, metaphor and meaning in Joni Mitchell's Blue. DUO Research Archive. University of Oslo Library. http://urn.nb.no/URN:NBN:no-14237
29. Fitch, W. T. (2015). Four principles of bio-musicology. Philosophical Transactions of the Roy-

al Society B: Biological Sciences, 370(1664), 20140091.

30. Fitzhugh, M. C., Whitehead, P. S., Johnson, L., Cai, J. M., Baxter, L. C., & Rogalsky, C. (2019). A functional MRI investigation of crossmodal interference in an audiovisual Stroop task. PloS one, 14(1), e0210736.
31. Flick, U. (2002). An introduction to qualitative research. Sage.
32. Gibson, J. J. (1979). The ecological approach to visual perception. Hillsdale, NJ: Lawrence Erlbaum Associates.
33. Gibson, E. J. (1989). Learning to perceive or Perceiving to learn? International Society for Ecological Psychology, Oxford, OH.
34. Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. Annual Review of Psychology 39, 1 – 41.
35. Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. Perception & psychophysics, 67(2), 274-284.
36. Gsöllpointner, K. (2015). Digital synaesthesia. The merge of perceiving and conceiving. Yearbook of Moving Image Studies. Vol. 1. http://filmbildtheorie.de/?page_id=577
37.
38. Hatfield, G. (2019). Gibson and Gestalt: (re)presentation, processing, and construction. Springer Nature B.V
39. Hargreaves, D.J. (2012). Imagination and creativity in music listening. Oxford University Press. UK.
40. Harrison, J., & Baron-Cohen, S. (1996). Acquired and inherited forms of cross-modal correspondence. Neurocase Vol. 2, pp. 245-249. Oxford University Press.
41. Herbert, R. (2013). Everyday music listening: Behaviour. Absorption, dissociation and trancing. Ashgate. Publishing, Ltd.
42. Herriott, R. & Mok, J. (2021). Problem Solving in Design and Music. Improvisation and Creation. Wassard Elea Rivista 2021. VII, no.2. WER no. 33. Ascena, Italy. p.74-97.
43. Holland, S., Wilkie, K., Mulholland, P., & Seago, A. (2013). Music interaction: understanding music and human-computer interaction. In Music and human-computer interaction (pp. 1-28). Springer, London.
44. Holland, S., McPherson, A. P., Mackay, W. E., Wanderley, M. M., Gurevich, M. D., Mudd, T. W., ... & Johnston, A. (2016, May). Music and HCI. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 3339-3346).
45. Holland, S., Mudd, T., Wilkie-McKenna, K., McPherson, A., & Wanderley, M. M. (2019). Understanding music interaction, and why it matters. In New directions in music and human-computer interaction (pp. 1-20). Springer, Cham.
46. Høffding, S., & Schiavio, A. (2021). Exploratory expertise and the dual intentionality of music-making. Phenomenology and the Cognitive Sciences, 20(5), 811-829.
47. Huron, D. (2015). Cues and signals: an ethological approach to music-related emotion. Signata 6, 331–351. doi: 10.4000/signata.1115
48. Juslin, P. N., & Sloboda, J. A. (2001). Music and

emotion. Theory and research.

49. Kawakami, A., Furukawa, K., Katahira, K., & Okanoya, K. (2013). Sad music induces pleasant emotion. Frontiers in psychology, 4, 311.

50. Klopfer, D. S. (1991). Apparent reversals of a rotating mask: A new demonstration of cognition in perception. Perception & psychophysics, 49(6), 522-530.

51. Köhler, W. (1929). Gestalt psychology. Liveright. New York.

52. Köhler, W. (1947). Gestalt psychology (2nd Ed.). Liveright. New York.

53. Kortum, P. (2008). HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces. Elsevier.

54. Lacey, S., Martinez, M., McCormick, K., & Sathian, K. (2016). Synaesthesia strengthens sound-symbolic cross-modal correspondences. European Journal of Neuroscience, 44(9), 2716-2721.

55. Langlois, T., Schloss, K., & Palmer, S. (2013). Music-Colour Associations to Simple Melodies in Synesthetes and Non-synesthetes. Journal of Vision, 13(9), 1325-1325.

56. Lakoff, G., & Johnson, M. (1980). Metaphors we live by. University of Chicago press. ISBN: 978-0226468013

57. Lakoff, G. & Johnsen, M. (2003) Metaphors we live by. The university of Chicago press. London.

58. Lin, A., Scheller, M., Feng, F., Proulx, M.J. & Metatla, O. (2021). Feeling colours: crossmodal correspondences between tangible 3D objects, colours and emotions. CHI Conference on Human Factors in Computing Systems. ACM, New York, USA.

59. Linson, A., & Clarke, E. F. (2017). Distributed cognition, ecological theory, and group improvisation. Distributed creativity: Collaboration and improvisation in contemporary music, 2, 52.

60. Lindborg, P., & Friberg, A. K. (2015). Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews. PloS one, 10(12), e0144013.

61. McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2021). Neural basis of the sound-symbolic crossmodal correspondence between auditory pseudowords and visual shapes. Multisensory Research, 35(1), 29-78.

62. McDonald, J., Canazza, S., Chmiel, A., De Poli, G., Houbert, E., Murari, M., ... & Zhang, J. D. (2022). Illuminating Music: Impact of Color Hue for Background Lighting on Emotional Arousal in Piano Performance Videos. Frontiers in psychology, 13, 828699-828699.

63. McDonnell, M. (2020a). Constructing visual music images with electroacoustic music concepts in, sound and image: audiovisual practice and aesthetics, Routledge.

64. McDonnell, M. (2020b). 'Digital Alchemy' - Sound and Image: Aesthetics and Practices, SeenSound Livestream. http://www.seeingsound.co.uk/seeing-sound-2018/2018-screenings/

65. Mencke, I., Omigie, D., Wald-Fuhrmann, M., & Brattico, E. (2019). Atonal music: Can uncertainty lead to pleasure?. Frontiers in Neuroscience, 12, 979.

66. Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. Neuroreport, 12(11), 2557-2560.

67. Mok, J. (2017). Alche(memory)- A memory game based on the neurological phenomenon, Synaestehsia. Universal design and digital UX/UI design categories. IDI Graduate Design Awards. Dublin, Ireland.
68. Mok, J. (2018). Advancing web accessibility to be more dyslexia-friendly. Inclusion Workshop. NordiCHI conference. University of Oslo, Norway.
69. Miles, M. B., Huberman, A. M., & Saladña, J. (2020). Qualitative data analysis. A methods sourcebook. Fourth Edition. SAGE. USA.
70. MIRAGE. (2022). https://www.uio.no/ritmo/english/projects/mirage/ (accessed January 10, 2022.)
71. Music Lab. (2021). MusicLab 8: Synaesthesia. Science Library, Oslo. https://www.uio.no/ritmo/english/news-and-events/events/musiclab/2021/musiclab-8-synaesthesia/musiclab-8-synaesthesia.html
72. (accessed Jan 10, 2022.)
73. Mencke, I., Omigie, D., Wald-Fuhrmann, M., & Brattico, E. (2019). Atonal music: Can uncertainty lead to pleasure?. Frontiers in Neuroscience, 12, 979.
74. Narmour, E. (2000). Music expectation by cognitive rule-mapping. Music Perception, 17(3), 329-398.
75. Norman, D. A. (1988). The psychology of everyday things. Basic books.
76. Olivetti Belardinelli, M., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., ... & Romani, G. L. (2004). Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. Cognitive Processing, 5(3), 167-174.
77. Obrist, M., Gatti, E., Maggioni, E., Vi, C. T., & Velasco, C. (2017a). Multisensory experiences in HCI. IEEE MultiMedia, 24(2), 9-13.
78. Obrist, M., Ranasinghe, N., & Spence, C. (2017b). Multisensory human–computer interaction. International Journal of Human-Computer Studies, 107.
79. Omigie, D., & Ricci, J. (2022). Accounting for expressions of curiosity and enjoyment during music listening. Psychology of Aesthetics, Creativity, and the Arts.
80. Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. Proceedings of the National Academy of Sciences, 110(22), 8836-8841.
81. Palmer, S. E., Langlois, T. A., & Schloss, K. B. (2016). Music-to-colour associations of single-line piano melodies in non-synesthetes. Multisensory Research, 29(1-3), 157-193.
82. Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. Proceedings of the National Academy of Sciences, 111(16), 6104-6108.
83. Patton, M. Q. (2014). Qualitative research & evaluation methods: Integrating theory and practice. Sage publications.
84. Pravossoudovitch, K., Cury, F., Young, S. G., & Elliot, A. J. (2014). Is red the colour of danger? Testing an implicit red–danger association. Ergonomics, 57(4), 503-510.
85. Ramachandran, V.S., & Hubbard, E.M. (2001). Psychophysical investigations into the neural basis of synaesthesia. Proceedings of the Royal So-

ceity B: Biological Sciences, 268 (1470), 979-983.

86. Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. Brain and Language, 128(1), 18-24.
87. Rothen, N., Tsakanikos, E., Meier, B., & Ward, J. (2013). Coloured Letters and Numbers (CLaN): a reliable factor-analysis based synaesthesia questionnaire. Consciousness and cognition, 22(3), 1047-1060.
88. Rudenko, S., & Cabral, J. P. (2019). Synaesthesia: How can it be used to enhance the audio-visual perception of music and multisensory design in digitally enhanced environments?. In Proc. SMM19, Workshop on Speech, Music and Mind 2019 (pp. 60-64).
89. Ruud, E. (1998). Music Therapy: Improvisation, Communication, and Culture. Barcelona Publishers. Spain.
90. Salice, A., Høffding, S., & Gallagher, S. (2019). Putting plural self-awareness into practice: the phenomenology of expert musicianship. Topoi, 38(1), 197-209.
91. Schmitz, L., Knoblich, G., Deroy, O., & Vesper, C. (2021). Crossmodal correspondences as common ground for joint action. Acta Psychologica, 212, 103222.
92. Schubert, E., Murari, M., Rodà, A., Canazza, S., Da Pos, O., & De Poli, G. (2019). Verbal and Cross-Modal Ratings of Music: Validation and Application of an Icon-Based Rating Scale. i-Perception, 10(3), 2041669519852643.
93. Sutcliffe, A. (2003). Multimedia and virtual reality: designing multisensory user interfaces. Psychology Press.
94. Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. Physics of life reviews, 7(3), 269-284.
95. Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.
96. Sloboda, J. A., Lamont, A., & Greasley, A. (2009). Choosing to hear music. The Oxford handbook of music psychology, 1, 431-440.
97. Spence, C. (2011). Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics, 73(4), 971-995.
98. Spence, C., & Parise, C. V. (2012). The cognitive neuroscience of crossmodal correspondences. i-Perception, 3(7), 410-412.
99. Spence, C., & Deroy, O. (2013). Crossmodal mental imagery. Multisensory imagery, 157-183.
100. Spence, C. (2020a). Simple and complex crossmodal correspondences involving audition. Acoustical Science and Technology, 41(1), 6-12.
101. Spence, C. (2020b). Senses of place: architectural design for the multisensory mind. Cognitive Research: Principles and Implications, 5(1), 1-26.
102. Spence, C. (2021). Sensehacking. Viking Penguin.
103. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18(6), 643–662. https://doi.org/10.1037/h0054651
104. Schiavio, A., Moran, N., Van der Schyff, D., Biasutti, M., & Parncutt, M. (2020). Processes and experiences of creative cognition in seven Western classical composers. Musicae Scientiae,

1-23, SAGE.

105. Szokolszky, A., Read, C., Palatinus, Z., & Palatinus, K. (2019). Ecological approaches to perceptual learning: learning to perceive and perceiving as learning. Adaptive Behavior, 27(6), 363-388.

106. Temperley, D. (2001). The Cognition of Basic Musical Structures. Musicae Scientiae, 6(1).

107. Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. Neuropsychologia, 63, 43-50.

108. Thaut, M. H. (2015). Music as therapy in early history. Progress in brain research, 217, 143-158.

109. Ullman, S. (1980). Against direct perception. Behavioral and Brain Sciences, 3(3), 373-381.

110. Umeozor, S. N. (2020). Information Retrieval: A Communication Process in the 21st Century Library. International Journal of Knowledge Content Development & Technology, 10(2), 7-18.

111. Van Goethem, A., & Sloboda, J. (2011). The functions of music for affect regulation. Musicae scientiae, 15(2), 208-228.

112. Van Der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory. Music & Science, 1, 2059204318792319.

113. Velasco, C., Woods, A.T., Deroy, O., & Spence, C. (2015). Hedonic mediation of the crossmodal correspondence between taste and shape. Food quality and preference 41. 151-158. Elsevier.

114. Vi, C. T., Ablart, D., Arthur, D., & Obrist, M. (2017, November). Gustatory interface: the challenges of 'how'to stimulate the sense of taste. In Proceedings of the 2nd ACM SIGCHI International Workshop on Multisensory Approaches to Human-Food Interaction (pp. 29-33).

115. Wade, N., & Swanston, M. (2013). Visual perception: An introduction. Psychology Press.

116. Watson, M. R., Akins, K., Spiker, C., Crawford, L., & Enns, J. T. (2014). Synaesthesia and learning: a critical review and novel theory. Frontiers in Human Neuroscience, 8, 98.

117. Zbikowski, L.M. (2012). Metaphor in nonverbal expression. Chapter. 28. Metaphor and music. Cambridge University press.

Appendices

Appendix 1 - Survey questions (Phase 1 and 2b):
https://doi.org/10.5281/zenodo.6552304

Appendix 2 - Interview protocols:

Protocol (Performer):
https://doi.org/10.5281/zenodo.6552304

Protocol (Phase 2a):

https://doi.org/10.5281/zenodo.6552304

Appendix 3 - A/V recording (concert setting):

M1:

https://youtu.be/6ESEvxYbQoA

M2:

https://youtu.be/za0xaeNnT30

Appendix 4 - Conceptual codes (Clusters):
https://doi.org/10.5281/zenodo.6552304

Appendix 5 - Survey responses (Phase 2b):
https://doi.org/10.5281/zenodo.6552304

Appendix 6 - Survey responses (Phase 1):

https://doi.org/10.5281/zenodo.6552304

Appendix 7 - Interactive materials (non-concert setting):

M1 (App):

This link is a test version for Apple IOS devices (copyright: MIRAGE): https://testflight.apple.com/join/fvh8t5Fh

Here is a quick demonstration of the app: https://youtu.be/21DyoZRp3wo

M2 (Video):

https://youtu.be/MVy4Bm97xkM

Appendix 8 - MCT Blog post:

https://mct-master.github.io/masters-thesis/2022/05/12/jonimok-crossmodal-correspondences.html